DOCUMENT RESUME

ED 327 212 IR 053 394

AUTHOR Lynn, M. Stuart
TITLE Preservation and Access Technology. The Relationship between Digital and Other Media Conversion Processes: A Structured Glossary of Technical Terms.
INSTITUTION Commission on Preservation and Access, Washington, DC.
PUB DATE Aug 90
NOTE 76p.; A Report of the Technology Assessment Advisory Committee of the Commission on Preservation and Access.
AVAILABLE FROM Commission on Preservation and Access, 1785 Massachusetts Avenue, NW, Suite 313, Washington, DC 20036 ($5.00 per copy prepaid).
PUB TYPE Reference Materials - Vocabularies/Classifications/Dictionaries (134)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Definitions; Digital Computers; Information Retrieval; *Information Storage; Library Automation; *Media Adaptation; Microforms; Paper (Material); *Preservation; Technological Literacy

ABSTRACT

This document presents a comprehensive glossary of terms which are associated with document preservation technologies; there is a particular emphasis on the technologies of media conversion and the use of digital computer technologies. The glossary also includes technologies associated with access to such preserved materials. A document such as this is designed to serve a duality of purposes: first, to provide clarity to a growing technical terminology, and second, to examine the impact of digital technologies on libraries. The overall structure of the glossary is organized by three major concepts: (1) the original document (medium, format, periodicity, properties, condition, content); (2) the selection process (by title, category, bibliography, use, condition, scholarly advisory committee, conspectus); and (3) the preserved copy (technologies of preservation and media conversion, capture, storage, access methodology, distribution, presentation). A subject index and 12 references are provided. (MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE COMMISSION ON PRESERVATION AND ACCESS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

REPORT

PRESERVATION AND ACCESS TECHNOLOGY

THE RELATIONSHIP BETWEEN DIGITAL AND OTHER MEDIA CONVERSION PROCESSES:

A STRUCTURED GLOSSARY OF TECHNICAL TERMS

ED327212

by

M. STUART LYNN

and

The Technology Assessment Advisory Committee
to the Commission on Preservation and Access

August 1990

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Maxine K. Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1785 Massachusetts Avenue, N.W., Suite 313, Washington, D.C. 20036 • (202) 483-7474

IR053394

The Commission on Preservation and Access was established in 1986 to foster and support collaboration among libraries and allied organizations in order to ensure the preservation of the published and documentary record in all formats and to provide enhanced access to scholarly information.

The Commission is accessible via ALANET (ALA2624) and BITNET (CPA@GWUVM, SITTS@GWUVM), and by FAX (202) 483-6410.

This document was printed by Cornell University with the same technology that is being used for printing books in connection with Cornell's digital book preservation study. This study is being supported in part by the Commission on Preservation and Access and in part by the Xerox Corporation. Since this printing technology is restricted to black-and-white, an exception has been made to the Commission's normal practice of using blue ink.

Published by
The Commission on Preservation and Access
1785 Massachusetts Avenue, NW, Suite 313
Washington, DC 20036

August, 1990

Additional copies available for $5.00 from the above address. Order must be prepaid, with check made payable to "The Commission on Preservation and Access." Payments must be in U.S. funds. Please do not send cash.

This publication has been submitted to
the ERIC Clearinghouse on Information Resources.

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984

Copyright © 1990 by the Commission on Preservation and Access. Copying without fee is permitted provided that copies are not made or distributed for direct commercial advantage and credit to the source is given. Abstracting with credit is permitted. To copy otherwise, or to republish, requires a fee and specific permission.

COMMITTEE PREFACE

In 1989, the Technology Assessment Advisory Committee (TAAC) of the Commission on Preservation and Access was asked by the Commission to consider the potentials of various new technologies for the capture of printed and other information now at risk, and the storage and retrieval of preserved materials. This report is one in a series alerting the Commission and others to developments and possibilities within the context of national and international initiatives for preservation of and access to information printed on disintegrated paper and other substrates. During its first meetings, the Committee found the need for a framework within which to discuss the use of emerging technologies for preservation purposes — a framework that could also be shared with professionals working in the preservation and related fields.

The resulting "structured glossary", which represents the views and thinking of the full TAAC membership, was principally authored by M. Stuart Lynn with assistance from colleagues in the libraries and information technologies divisions at Cornell University. This paper has also been subjected to a pre-publication review by selected members of the library and information technologies professions at large. The Committee hopes that this Glossary will contribute to a common understanding of how preservation and access needs can be addressed by emerging technologies, in order to take full advantage of appropriate opportunities.

Rowland Brown, Chair
Technology Assessment Advisory Committee

TAAC membership consists of representatives of the computer and communications industries, as well as corporate and higher education institutional consumers of advanced technologies. The members are: Adam Hodgkin, Managing Director, Cherwell Scientific Publishing Limited; Douglas van Houweling, Vice Provost for Information Technologies, University of Michigan; Michael Lesk, Division Manager, Computer Sciences Research, Bellcore; M. Stuart Lynn, Vice President for Information Technologies, Cornell University; Robert Spinrad, Director, Corporate Technology, Xerox Corporation; Robert L. Street, Vice President for Information Resources, Stanford University; and Rowland C.W. Brown, Chair, President, OCLC (retired).

ACKNOWLEDGEMENTS

The Committee is particularly grateful to John Dean, Conservation Librarian, Cornell University Library; and to Lynne K. Personius, Assistant Director for Scholarly Information Technologies, Cornell Information Technologies, for their assistance in the preparation of this paper. The Committee also owes a special debt of gratitude for their careful review of the paper to Margaret Byrnes, Head Preservation Section, National Library of Medicine, and to Gay Walker, Head Librarian for Preservation, Yale University Library. Invaluable additional comments were also provided by Millicent Abell, University Librarian, Yale University; Richard De Gennaro, Roy E. Larsen Librarian, Harvard University; James F. Govan, University Librarian, University of North Carolina at Chapel Hill; Paula Kaufman, Dean of Libraries, University of Tennessee; and Michael Keller, Associate University Librarian for Collections Development, Yale University.

FOREWORD

This document is offered as a structured glossary of terms associated with the technologies of document *preservation*, with particular emphasis on document *media conversion* technologies (often called "reformatting technologies"), and even more particularly on the use of *digital computer* technologies. The Glossary also considers technologies associated with access to such preserved documents. Such a glossary is intended for communication among people of different professional backgrounds, especially since in recent years there has been a proliferation of such technologies and associated technical terms, technologies and terms that cut across many disciplines.

The use of digital technologies, however, has implications for libraries that extend far beyond the boundaries of preservation and of access to preserved materials. Some of these implications are summarized in the discussion in the Introduction of "The Impact of Digital Technologies," and are indicated throughout the Glossary. Thus this Glossary may serve a wider purpose than the title itself would imply.

The Glossary is a *structured* glossary, in the sense that the defined terms have been hierarchically grouped. The term "taxonomy" was used to describe earlier drafts of the manuscript, but that term was dropped since it might imply a degree of completeness and form beyond that envisaged, or even possible. The Glossary is not intended to be complete with respect to preservation technology as a whole, but is highly selective (and even highly subjective) in its choice of terms to include, and very much slanted towards the use and impact of digital technologies. Other preservation technologies are sketched in for contextual purposes only. Within these constraints, the Glossary is intended to be comprehensive but not exhaustive.

The Glossary is not intended to be so comprehensive as to satisfy the technologist only concerned with technologies, or the librarian exclusively concerned with librarianship and preservation. It is intended to satisfy the intersection of their concerns. On the other hand, issues of preservation and access raise concepts that have implications for librarianship as a whole, so that, in that sense, this Glossary has consequences that are not limited to the preservation arena alone.

PRESERVATION AND ACCESS TECHNOLOGY

THE RELATIONSHIP BETWEEN DIGITAL AND OTHER MEDIA CONVERSION PROCESSES: A STRUCTURED GLOSSARY OF TECHNICAL TERMS

TABLE OF CONTENTS

COMMITTEE PREFACE i

ACKNOWLEDGEMENTS ii

FOREWORD iii

TABLE OF CONTENTS v

INTRODUCTION 1

The Impact of Digital Technologies 1

Scope of the Glossary 5

Structure of the Glossary 7

1. THE ORIGINAL DOCUMENT 9

1.1. Document Medium 10
1.1.1. Paper 10
1.1.2. Microform 11
1.1.3. Video 11
1.1.4. Film 12
1.1.5. Audio 12
1.1.6. Digital Electronic 12
1.1.6.1 Magnetic Disk 13
1.1.6.2 Magnetic Tape 13
1.1.6.3 Optical Disk 13
1.1.6.4 Optical Tape 13
1.1.6.5 Magneto-Optical Disk 13
1.1.7. Multi-Media 13

1.2. Document Format 14
1.2.1. Manuscript 15
1.2.2. Book 15
1.2.3. Pamphlet 15
1.2.4. Newspaper 15
1.2.5. Printed Sheet 15

1.2.6. Periodical 15
1.2.7. Cartographic Materials 15
1.2.8. Music 16
1.2.9. Graphic Materials 16
1.2.9.1 Art Original 16
1.2.9.2 Filmstrip 16
1.2.9.3 Photograph 16
1.2.9.4 Picture 16
1.2.9.5 Technical Drawing 16
1.2.9.6 Miscellaneous 16
1.2.10. Data File 16
1.2.10.1 Table 17

1.3. Document Periodicity 17
1.3.1. Monograph 17
1.3.2. Serial 17

1.4. Document Properties 18
1.4.1. Tone 18
1.4.1.1 Monotone 18
1.4.1.1.1 Two-Tone 18
1.4.1.1.2 Greyscale 18
1.4.1.2 Highlight Color 19
1.4.1.3 Two Color 19
1.4.1.4 Full Color 19
1.4.2. Object Type 19
1.4.2.1 Text Objects 19
1.4.2.2 Data Objects 19
1.4.2.2.3 Table 19
1.4.2.3 Graphic Objects 19
1.4.2.3.1 Line Art 20
1.4.2.3.1.1 Graphs 20
1.4.2.3.2 Halftone 20
1.4.2.3.3 Discrete Tone 20
1.4.2.3.4 Continuous Tone 20

1.5. Document Condition 20
1.5.1. Archival 21
1.5.2. Non-Archival 21
1.5.3. Acidic 21
1.5.4. Brittle 21
1.5.5. Other 22

1.6. Document Content 22
1.6.1. Intellectual Content 22
1.6.2. Copyright 22
1.6.3. Structure 23
1.6.3.1 Abstract (see 3.4.1.2) 24
1.6.3.2 Title Page 24
1.6.3.3 Table of Contents (see 3.4.1.3) 24
1.6.3.4 List of Figures, Tables, Maps or Other Illustrations 24
1.6.3.5 Preface (see 3.4.1.5) 24
1.6.3.6 Introduction (see 3.4.1.6) 24
1.6.3.7 Body 24
1.6.3.8 Index (see 3.4.1.7) 24
1.6.3.9 Other 24

2. THE SELECTION PROCESS 25

2.1. By Title 26

2.2. By Category 26

2.3 By Bibliography 27

2.4. By Use 27

2.5. By Condition 27

2.6. By Scholarly Advisory Committee 27

2.7. By Conspectus 27

3. THE PRESERVED COPY 29

3.1. Preservation and Media Conversion Technologies 30
3.1.1. Conservation Treatment 31
3.1.2. Paper Deacidification and Strengthening 31
3.1.3. Photocopying 32
3.1.4. Microform Recording 33
3.1.5. Electronic Digitization 34
3.1.5.1 Image Document 36
3.1.5.2 Text Document 37
3.1.5.2.1 Unformatted Text 37
3.1.5.2.2 Formatted Text 37
3.1.5.3 Compound Document 37
3.1.6. Rekeying of Text 37
3.1.6.1 Unformatted Text 38
3.1.6.2 Formatted Text 38
3.1.7. Reprinting or Republication 38

3.2. Capture Technology 38
3.2.1. Photocopier 39
3.2.2. Microform Recorder 39
3.2.3. Digital Image Scanner 39
3.2.4. Optical Character Recognition Scanner 40
3.2.5. Internal Character Recognition 41
3.2.6. Intelligent Character Recognition 42
3.2.7. Page Recognition 42
3.2.8. Rekeying of Text 42
3.2.9. Enhancement 42

3.3. Storage Technology 43
3.3.1. Storage Medium 43
3.3.1.1 Paper (see 1.1.1) 43
3.3.1.2 Microform (see 1.1.2) 43
3.3.1.3 Video (see 1.1.3) 43
3.3.1.4 Film (see 1.1.4) 43
3.3.1.5 Audio (see 1.1.5) 43
3.3.1.6 Digital Electronic 43
3.3.1.6.1 Magnetic Disk 44
3.3.1.6.2 Magnetic Tape 45
3.3.1.6.3 Optical Disk 45
3.3.1.6.4 Optical Tape 45
3.3.1.6.5 Magneto-Optical Disk 46
3.3.2. Compression 46

3.3.2.1 Uncompressed 46
3.3.2.2 Reversibly Compressed 47
3.3.2.2.1 CCITT Group Compression 47
3.3.2.2.2 Reversible Textual Compression 47
3.3.2.2.3 Page Description Language Compression (PDL) 47
3.3.2.2.4 Other Compression Standards or Algorithms 47
3.3.2.3 Irreversibly Compressed 47
3.3.2.3.1 Irreversible Textual Compression 47
3.3.3. Storage Format 47
3.3.4. Encoding Method 48
3.3.4.1 No Encoding 48
3.3.4.2 Textual Encoding 48
3.3.4.3 Markup Language Encoding 49
3.3.4.4 Page Description Language Encoding 49
3.3.5. Useful Life 49

3.4. Access Methodology or Technology 50
3.4.1. Indexed Access 51
3.4.1.1 Via Catalog 51
3.4.1.2 Via Abstract 51
3.4.1.3 Via Table of Contents 51
3.4.1.4 Via List of Figures, Tables, Maps or Other Illustrations 52
3.4.1.5 Via Preface 52
3.4.1.6 Via Introduction 52
3.4.1.7 Via Index 52
3.4.1.8 Via Citation 52
3.4.2. Full (or Partial) Document Access 52
3.4.2.1 Via Inverted Text File Index 53
3.4.3. Compound Document Access 53

3.5. Distribution Technology 54
3.5.1. Distribution Medium 54
3.5.1.1 Paper (see 1.1.1) 54
3.5.1.2 Microform (see 1.1.2) 54
3.5.1.3 Video (see 1.1.3) 54
3.5.1.4 Film (see 1.1.4) 54
3.5.1.5 Audio (see 1.1.5) 54
3.5.1.6 Digital Electronic (see 1.1.6) 54
3.5.2. Messenger Services 55
3.5.3. FAX 55
3.5.4. Print-on-Demand 55
3.5.5. Data Networks 56
3.5.5.1 Local Area Network 56
3.5.5.2 Wide Area Network 56
3.5.5.3 National Network 57
3.5.6. Voice Networks 57
3.5.7. Cable Networks 57

¡3.6. Presentation Technology 57
3.6.1. Presentation Medium 58
3.6.1.1 Paper (see 1.1.1) 58
3.6.1.2 Microform (see 1.1.2) 58

3.6.1.3 Video (see 1.1.3) 58
3.6.1.4 Film (see 1.1.4) 58
3.6.1.5 Audio (see 1.1.5) 58
3.6.1.6 Digital Electronic (see 1.1.6) 58
3.6.2. Presentation or Viewing Device 58
3.6.2.1 Paper Document 58
3.6.2.2 Microform Reader 58
3.6.2.3 Video Projector (Television Set) 58
3.6.2.4 Film, Slide, or Other Projectors 59
3.6.2.5 Audio Devices 59
3.6.2.6 Computer Workstation 59
3.6.2.6.1 Display Monitor 59
3.6.2.6.2 Local Printer 60
3.6.2.6.3 Remote Printer 60
3.6.2.6.4 Other Local Media Output Device 60
3.6.2.7 Multi-Media Workstation 60

4. SOURCES OF INFORMATION 61

INDEX 63

PRESERVATION AND ACCESS TECHNOLOGY

THE RELATIONSHIP BETWEEN DIGITAL AND OTHER MEDIA CONVERSION PROCESSES: A STRUCTURED GLOSSARY OF TECHNICAL TERMS

A Report Of
The Technology Assessment Advisory Committee
of
The Commission on Preservation and Access

INTRODUCTION

This document is offered as a structured glossary of terms associated with the technologies of document *preservation*, with particular emphasis on document *media conversion* technologies (often called "reformatting technologies"),[1] and even more particularly on the use of *digital computer* technologies. The Glossary also considers technologies associated with access to such preserved documents. Such a glossary is intended for communication among people of different professional backgrounds, especially since in recent years there has been a proliferation of such technologies and associated technical terms, technologies and terms that cut across many disciplines.

The use of digital technologies, however, has implications for libraries that extend far beyond the boundaries of preservation and of access to preserved materials. Some of these implications are summarized in the following discussion of "The Impact of Digital Technologies," and are indicated throughout the Glossary. Thus this Glossary may serve a wider purpose than the title itself would imply.

The Impact of Digital Technologies

The digital computer technology revolution continues to open up concepts, many of which are only just beginning to be understood or accepted. These concepts are critically important to librarianship in general and preservation in particular. In a world historically dominated by paper, the same medium is used for document capture (creation, recording),

1 See Section 3.1 for a discussion of the use of the term "media conversion" to replace the use of the term "reformatting." We also follow the distinction that while media conversion is not a *conserving* technology, it is a *preserving* technology.

storage, access, distribution and use, and there has been no compelling need to consider these as separate entities. There has also been no compelling need to distinguish between the *format* of a document and the *medium* in which it is embodied, since there is only one dominant choice of medium. Indeed, the terms have traditionally been used somewhat interchangeably and indiscriminately. The introduction of non-paper forms such as phonograph recordings and films has modified this straightforward view somewhat, but traditional cataloging makes every effort to foster the constraint that there is a one-to-one correspondence between the format and the medium, with the objective of identifying the combined format-medium with some physical shelf location.

Further efforts to foster this constraint increasingly break down when digital technologies enter the picture. Digital technologies open a world that paradoxically is simultaneously more complex and, in some ways, simpler. It is more complex because now the same document or document format may intrinsically be represented in different media for different purposes, forcefully motivating the need to distinguish carefully between the format and the medium. Furthermore, different media may be used interchangeably for different stages of document handling, that is, for capture, storage, access, distribution, and use. To complicate the situation even more, the documents may be encoded in a myriad of ways at each of these stages.

And yet, separation of the format and the medium — and treating each stage of document handling separately — may open up a more logical structure free from traditional constraints. In this sense, digital technologies may simplify certain aspects of librarianship.

Digital technologies present many new challenges, however, that must be considered. For example, although these varying formats may be decoded and translated back and forth among each other, many fear that the means of decoding may become lost as a result of technological obsolescence, conceivably making digitally stored documents inaccessible. There are also many who question the longevity of the physical media used in digital technologies. Others suggest that the appropriate way to address both of these problems — as well as to take advantage of the declining costs of computer storage and of increasing storage densities — may well be to copy stored documents periodically onto new media.

Indeed the main advantage of the world of digital technologies, namely that they represent a kind of "esperanto" of mutually comprehensible and interchangeable formats, may, if not properly managed, also represent their biggest weakness, because of the rapidity of change and obsolescence, and because of the wide range of choices available at any given time. Their

very attractiveness could lure the unwary or the uninformed into dangerous territory.

Periodic recopying onto new media represents a whole new approach for libraries to the operation and financing of "inventory management" (although though such practices are quite common in data centers). The implications could be quite extensive. Librarians tend to think in terms of periods of centuries rather than having (or wanting) to recopy every few years. Such considerations may either hinder the adoption of digital technologies for preservation or other purposes or eventually cause some rethinking of the underlying economics of librarianship.

The incentive for such potential reevaluation, however, is not limited to the preservation of older materials, nor is the influence of technology the only driving factor. The underlying stimulus is a gradual transition over the centuries — perhaps spurred by the exponential growth of recorded knowledge and information — from documents with associated physical or conceptually useful lifetimes, times between new editions, or, more generically, times between "instances", that can be measured in decades or centuries; to documents with associated times between instances measured in much shorter units of time — even, in the case of "active documents" (see below), measured in minutes or seconds.

In essence, this represents a transition from "batch processing" to "continuous processing."[2] The financial and other implications of this could undoubtedly be far-reaching for libraries (a full discussion is beyond the scope of this Glossary), introducing into the library milieu unfamiliar (or, at least, largely unused) concepts associated with continuous processes or processes with relatively short lifetimes, such as "depreciation" and "lifecycle costing." These are concepts that are familiar to the world of digital electronic processing and quite normal outside of universities, but that have been avoided in worlds — such as research libraries — that depend to a greater or lesser extent upon irregular gifts or grants of varying or unpredictable size, donations directed to the purchase and immediate storage of documents, but not to their maintenance. Indeed, one of the most serious questions facing librarians in the future may be how to effect a match between the changing economic demands of "continuous processes" and the traditional nature of many funding sources. Will donors, for example, be as willing to support the continuous demands of technological processing as they have historically and generously supported the periodic construction of library buildings? What

2 This analogy was pointed out by Douglas van Houweling.

implications does the financing of continuous processes have for the "free" and openly accessible library?[3]

Yet the potential of digital technologies and of the flexibility they offer is boundless. Over the coming decades, these technologies may open up vistas of ever-increasing storage densities to where entire libraries can be electronically stored in the space of a single room; of blinding access and distribution speeds allowing whole documents to be moved almost instantly across the nation's (and indeed the world's) data networks, leading to the concept of the "distributed library;" of ease of replication at very modest cost (another cause for alarm, particularly to those concerned with protection of intellectual property); of "print-on-demand" where paper copies of documents are only printed "just in time" and not inventoried in advance of need; of accessibility at a distance away from where the "digital document" or preservation copy was created or is stored; and of intelligent automated document analysis. Indeed, the means of creation and production of documents have already been revolutionized by these technologies.

These technologies also open up horizons for totally new document formats, such as *active* documents whose contents may combine different media such as text, sound, video or voice; or whose contents may change dynamically with time, what Harvey Wheeler called "the fungible book."[4] The preservation of these new "active" formats is not of direct interest to the subject of preservation of more traditional formats (and therefore beyond the scope of this Glossary), but is of indirect interest because digitally preserved traditional documents can be incorporated into such active documents. Furthermore, contemporary active documents will become a subject of future preservation interest.

Some view the introduction of digital technologies into the world of libraries as likely to cause a revolution as far-reaching as that caused by the printing-press: a massive paradigm shift. Others view the introduction with concern (one cannot help but recall that the monks at first also viewed the introduction of the printing press with equal concern), an intimidating perturbation that disturbs an equilibrium and modalities of scholarship that have served well for many decades or even for centuries.

Either way, digital technologies cannot be ignored. They are already with us. The question is not whether they will have a presence, but the pace

3 A glimpse of possible implications has already been seen in the tendency of many libraries to charge patrons for searches of electronic databases.

4 Harvey Wheeler: "The Virtual Library: The Electronic Library Developing Within The Traditional Library". Doheny Documents. University of Southern California University Library. 1987.

and degree to which that presence will grow and influence. The next twenty years are likely to be times of extraordinary change. Our libraries — indeed our universities, colleges, and our scholarly communities — may well be remade by the consequences of this technological revolution.

And yet — in spite of technology's impact and of the revolutionary consequences of that impact — it must be recognized that technology itself is not the ultimate driving force. It is the inexorable pressure caused by the exponential growth of recorded knowledge, and the ever-increasing complexity, costs, and other problems associated with the storage and distribution of, and access to, such information. Technology can provide some solutions: it is not an end in itself.

Furthermore — for many reasons too numerous to detail here — the "digital library" is not about to replace the "paper library." Both will need to coexist in a shifting environment, at least for the foreseeable future. This in itself will present librarians with many economic, organizational, social, technical, and other challenges.

Between the eager apostles of technology and those who approach change with extreme caution lies the mass of professionals who are trying to understand and grapple with the potential of this shifting environment, many of them implementing prototype activities designed to elucidate greater insight,[5] many working to close the gap between promise and reality.

It is to these professionals — from all fields — that this Glossary is dedicated, to provide a common language for dialogue and mutual understanding, particularly as is required to address the problems of preservation, and the potential application of digital technologies to those problems. The Glossary is not intended to be so comprehensive as to satisfy the technologist only concerned with technologies, or the librarian exclusively concerned with librarianship and preservation. It is intended to satisfy the intersection of their concerns. On the other hand, issues of preservation and access raise concepts that have implications for librarianship as a whole, so that, in that sense, this Glossary has consequences that are not limited to the preservation arena alone.

Scope of the Glossary

This document is a *structured* glossary, in the sense that the terms have been hierarchically grouped. The term "taxonomy" was used to describe earlier drafts of the manuscript, but that term was dropped since it might

5 Some fields, particularly those propelled by the impetus of commercial endeavors such as medicine, law, and finance, are beyond the prototype stage and are into full production.

imply a degree of completeness and form beyond that envisaged, or even possible, for such a document. This document is not intended to be complete with respect to preservation and access technologies as a whole, but is highly selective (and even highly subjective) in its choice of terms to include, and very much slanted towards the use and impact of digital technologies. Other preservation technologies are sketched in for contextual purposes only. Within these constraints, the Glossary is intended to be comprehensive but not exhaustive.

The Glossary is not intended to solve all issues associated with the definition of technological and other terms associated with preservation and access. It is a conceptual document. Not all terms are defined with equal precision; indeed, the degree of precision is largely directed by the extent to which it is necessary to distinguish among these terms. The Glossary is intended to be adequate to support further research and development on the subject. Indeed, one measure of success of the Glossary will be the extent to which it stimulates additional work in the field, including refinements of the Glossary itself.

For the conceptual reasons outlined above, the Glossary departs from many well-established norms. Furthermore, excluded in any detail are terms primarily associated with *conservation*, such as paper deacidification, where every effort is made to preserve the documents in their original physical form,[6] or hand conservation. The focus, as stated, is on preservation through *media conversion* (traditionally known as "reformatting", a term which we do not favor in this Glossary — see 3.1), where the objective is to preserve the intellectual content of the original document on some other medium, and also if desired to produce at some later stage a close physical facsimile of the original, at least to the extent allowed by the technology.

The focus is also for the most part on *paper* documents requiring preservation. These represent the principal (but not the only) area of national and international attention: paper documents have the longest history and exist in the greatest numbers. They are also in urgent need of preservation because of the "embrittlement" (see 1.5.4) caused by the high acid content of paper manufactured since the mid-nineteenth century and by improper storage environments. In the years to come, the focus may well shift to other media. There is already, for example, considerable attention paid to film preservation, and video recordings are already deteriorating at an alarming rate.

6 Conservation may allow for only partial preservation of the original document. The bindings, for example, may be replaced while the body of the document is conserved.

Different technologies are more or less suitable to preserve different classes of documents or for achieving different access or other objectives. One of the main applications intended for this Glossary is for the classification of ranges of activity that can be used to describe different investigations into preservation and access methodologies. The level of detail varies throughout the Glossary according to what we believe is necessary to make the Glossary most pertinent to this intended application.

Structure of the Glossary

The Glossary is divided into three main sections: the Original Document, the Selection Process, and the Preserved Copy. The latter is dealt with in the most detail; in turn it is divided into a number of subsections: the first defines the actual preservation or media conversion technologies that may be employed; and the remaining subsections are devoted to the various technologies employed in the different stages of preservation and access — capture, storage, access, distribution, and presentation.

The reader will observe that there is some repetition of discussion of certain concepts throughout the Glossary. This is intentionally introduced, since it is expected that most readers will not choose to read the Glossary from cover to cover.

The overall structure of the Glossary is presented in Figure 1.

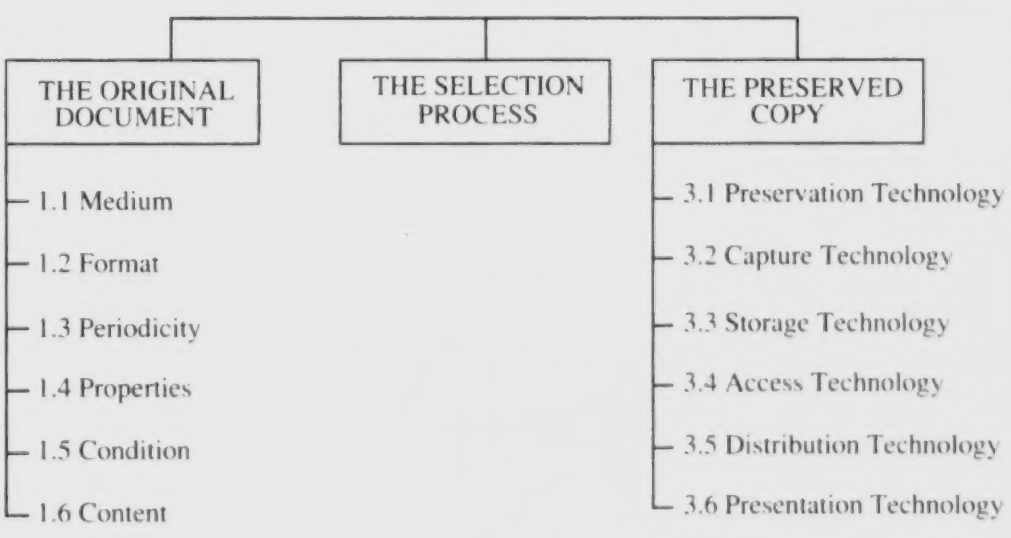
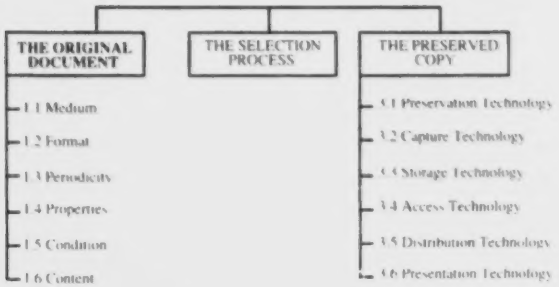


Figure 1: Overall Structure of Glossary



1. THE ORIGINAL DOCUMENT

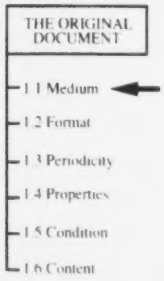
Different preservation or media conversion technologies are appropriate to different kinds of original material. This section, therefore, is devoted to a classification of terms used in describing the original document to be preserved, particularly those terms that need to be referenced in the context of media conversion.

The term *document* is used generically throughout this Glossary to include all forms of books, manuscripts, records and other classes of material containing information or other matter of intellectual content, regardless of the actual medium (1.1) or format (1.2) employed.

The Glossary takes free license with terms that have taken on a traditional meaning in the context of cataloging and other library activities, and in fact frequently departs from traditional norms used in this area. As stated in the Introduction, the reason for this is that such traditional definitions often confuse the *format* and *content* of the document with the *medium* used to record it, terms that have traditionally been used somewhat interchangeably and indiscriminately. This made sense when paper was the primary medium used for document capture, storage, distribution, and use. With newer technologies, however, and particularly with those used for media conversion (3.1), different media can be used for each of these stages, and, in fact, different media can be used for different instances of each stage. In this context, therefore, it makes taxonomic sense to separate format from medium.

For example, a traditional classification is "Motion pictures and video recordings." In our Glossary, the document format would be "motion pictures." The medium could be "film" or "videotape" or even "digital electronic" (such as with digital video). Even a book (document format) could be embodied in different media: "paper," "audio" (the "talking book"), "microform," or "digital electronic." To extend the example,

the book could be *stored* in a digital electronic medium, and subsequently *distributed* electronically, and *used* by "printing-on-demand" on paper or microform, or by presentation at a digital computer workstation.



1.1. Document Medium

***Document Medium* refers to the material upon which the original document was recorded.**

1.1.1. Paper

Paper is a medium traditionally used for printed books and other documents that are the most frequent target of preservation efforts. Paper is defined to be sheets usually made of vegetable fibers laid down on a fine screen from a water suspension. Marks are imprinted on the paper using any of a number of techniques including *handwriting or drawing* using a variety of media such as pencil, pen and ink, or pastel; *various forms of printing* using inks (numerous technologies are used to accomplish this); *photographic printing*, where paper coated with light-sensitive emulsion is exposed to various intensities of light); *xerographic printing*, where an electrically charged photoconductive insulating surface is selectively exposed to light and the latent image is developed with a resinous powder; *thermographic printing*, where the paper is exposed to a directed heat source that selectively modifies parts of the surface that may have been pre-treated with a heat-sensitive powder; and *chemical transfer printing*, where the surface of the paper is chemically coated and selectively modified by pressure or other means.

Parchment and *vellum* are not paper since they are made from the skins of sheep, goats, or calfskin.[7] We note them here for completeness.

Hard Copy is a term often used to denote any document produced on paper.

1.1.2. Microform

Microform refers to a document medium for producing or reproducing printed matter. It records *microimages*, that is, images too small to be read without some form of magnification. In a general sense, microforms may be on film (1.1.4) or paper (1.1.1), but for purposes of this Glossary the definition is restricted to film. Reading a microform requires the assistance of a microform reader (3.6.2.2). Microform comes in different styles including *microfilm* (a film roll that contains microimages arranged sequentially) and *microfiche* (sheets of film in which many microimages are arranged in a grid pattern). Both usually contain a header that can be read without magnification).

Microforms are an economic and compact form of document representation for archival storage, but are inconvenient to read when compared with a printed book. Microform technology is used as a preservation medium (3.1.4), as a means of saving space (such as for the convenient storage of newspapers) , or as a means of duplicating scarce or unique documents, that is, microreproductions of other original documents. However, microform is sometimes used for original documents, for example, those created on a computer and directly printed out onto a *computer-output-on-microfiche (COM)* device; and for microreproductions of material assembled for the purposes of releasing an original edition in microform.

1.1.3. Video

Video is normally an analog (see definition under 1.1.6) electronic technology for recording still or moving images, usually combined with sound (cf. 1.1.5). Following standards (which vary across the world) defined for television playback and broadcasting, the images are normally recorded on magnetic tape (3.3.1.6.2), when it is known as *videotape*, but also on other physical media such as optical disk (3.3.1.6.3) (*videodisk*).

7 Originally, the term "vellum" was restricted to calfskin. The distinction between parchment and vellum has eroded over the years.

Playback is usually achieved through a television set or video projector (3.6.2.3), although it is now possible and becoming common to play video recordings back through a computer (3.6.2.6) or multimedia workstation (3.6.2.7).

1.1.4. Film

Film is a recording medium consisting of thin sheets or strips of transparent or translucent material, such as polyester or acetate, coated with a light-sensitive emulsion. Recording occurs by exposing the film to the light emitted or reflected by the entity being recorded. Film is also the medium used for microfilm recording (1.1.2). A *photograph* (1.2.9.3) is produced using essentially the same technology, except that normally the light-sensitive emulsion is adhered to paper or some other opaque medium.

1.1.5. Audio

Audio documents are recordings made on a variety of (usually) magnetic media (see 3.3.1.6) of sounds only (as contrasted with video recordings (1.1.3) that also combine images). The evolution of such audio recordings has traversed a large number of different formats and physical media, including *phonograph disks (records)* of varying size (78 rpm's, 45 rpm's, 33 rpm's) and *tape cassettes* (of different formats), both of which are analog (see 1.1.6) recording technologies; and, more recently, *compact disks* and *digital acoustic tapes (DATs)*, which are digitally (1.1.6) encoded.

1.1.6. Digital Electronic

Digital Electronic Technologies[8] are technologies used to capture (3.2.3), store (3.3.1.6), transform (3.3.2, 3.3.4), distribute (3.5.1.6) or present (3.6.1.6, 3.6.2.6, 3.6.2.7) information in quantized electronic form (normally as a sequence of 0's and 1's known as *bits*). *Digital*, in which information is quantized discretely, is to be contrasted with *Analog*, in which information is not quantized but maintained in a continuous format.[9] A video

8 The term *digital technologies* is also used for brevity throughout this Glossary.

9 The non-technical reader may wish to compare the odometer of a car (a *digital* device which quantizes in precise 1/10th of a mile increments) with the speedometer (an *analog* device which displays speed continuously but which can only be interpreted approximately).

recording (1.1.3), is an example of an electronic technology that is analog[10].

For a variety of reasons, digital technologies are gradually replacing analog technologies. Reasons of importance to this Glossary are the convertibility of digital technologies among each other and into and from other technologies (such as paper and voice), so that digital technologies become a kind of *lingua franca* of communication and storage; and the ease of transmission of information by digital technologies across networks (3.5.5) to facilitate communication at a distance.

Original documents that are of concern for library preservation purposes are not normally encoded in a digital electronic medium.[11] Since this may become a subject of future concern, the category is included for completeness. Definitions, however, are more appropriately included under Storage Technology Medium (3.3.1.6).

1.1.6.1 Magnetic Disk (see 3.3.1.6.1)

1.1.6.2 Magnetic Tape (see 3.3.1.6.2)

1.1.6.3 Optical Disk (see 3.3.1.6.3)

1.1.6.4 Optical Tape (see 3.3.1.6.4)

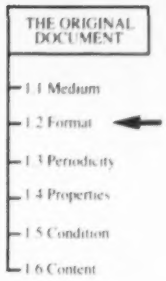
1.1.6.5 Magneto-Optical Disk (see 3.3.1.6.5)

1.1.7. Multi-Media

Multi-Media is a term used to denote documents created using a number of different media simultaneously, usually those with an electronic technological basis: for example, a digital electronic recording (1.1.6) that also combines video (1.1.3) and audio (1.1.5), and that may, as part of the document, intrinsically produce paper (1.1.1) outputs.

10 However, *digital*(ly-encoded) *video* is now becoming part of the panoply of technologies, where analog video signals are converted to digital signals for purposes of storage, transmission and playback through a computer (3.6.2.6) or multi-media (3.6.2.7) workstation.

11 This assertion, however, may not be true in the future. For example, music is now recorded in digital electronic form, such as DDD Compact Discs.



1.2. Document Format

Document Format* refers to the class of document with respect to its *style, arrangement,* or *layout.

Although this Glossary emphasizes the distinction between format and medium, some formats are more closely associated with a given medium. Thus, formats such as documentary, short, feature, and newsreel are most closely associated with the medium of film. Consistent with the main thrust of this Glossary, we emphasize those formats that are mostly associated with the medium of paper, even though several of these formats may also be embodied in other media (the "talking book," for example, recorded, say, on tape cassettes).

The term "format" itself may be too all-encompassing. There may be a need to further distinguish between the "type" of a document, such as "book," and the arrangement or layout of the book — such as formatted text on pages, or simply linear text that is not formatted into pages (as in the "talking book" where pages are not distinguished). However, this Glossary does not make this distinction, partly because of its focus on the paper milieu, where such a distinction may not be necessary, and partly because in the emerging world of digital technologies it may be premature to attempt such a distinction.

The use of the term "format" should not be confused with its use in the context of "reformatting." The latter, as described in 3.1, is best replaced by the term "media conversion."

1.2.1. Manuscript

For purposes of this Glossary, an original, unpublished document directly created by its author(s), usually on paper or parchment, and often in the author's own hand.

1.2.2. Book

A monograph (1.3.1) publication containing more than 49 pages, usually on paper.[12]

1.2.3. Pamphlet

A complete monograph (1.3.1) of at least 5 but not more than 49 pages, usually on paper (see Footnote 12).

1.2.4. Newspaper

A serial (1.3.2) publication issued at stated, frequent intervals containing news, opinions, advertisements, and other topical material, usually on paper (see Footnote 12).

1.2.5. Printed Sheet

A single sheet of printed paper such as a poster (but see 1.2.9.4), broadside, folded leaflet, or memorandum, usually on paper.

1.2.6. Periodical

A serial publication (1.3.2) appearing at regular or stated intervals, generally more frequently than annually, usually on paper (see Footnote 12). Includes *magazines* and *journals*.

1.2.7. Cartographic Materials

Representations of a selection of abstract features of the universe, most often in relation to the surface of the earth, often on paper but also on other substrates.

12 Although an increasing number of books are published on other media (see the Introduction to this Section). This remark also applies to 1.2.3, 1.2.4, 1.2.5, 1.2.6, and 1.2.8. Video magazines and journals, for example, are beginning to appear. A few books are being published *only* in digital form for playback on a computer workstation.

1.2.8. Music

In this context, printed representation of musical notation for instrumental, chamber, orchestral, and vocal scores, usually on paper (see footnote 12).

1.2.9. Graphic Materials

1.2.9.1 Art Originals, Prints, and Reproductions

Illustrated works, such as drawings, engravings, and lithographs, issued separately from books.

The following terms are included for completeness, but without definition[13]:

1.2.9.2 Filmstrips
1.2.9.3 Photographs, Slides, Transparencies, and Stereographs
1.2.9.4 Pictures, Postcards, and Posters
1.2.9.5 Technical Drawings (including Architectural Plans)
1.2.9.6 Miscellaneous

The *Miscellaneous* category includes flash cards, radiographs, study prints, and wall charts.

1.2.10. Data File

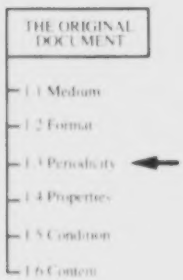
The term *Data File* is used generically to denote a document consisting of a collection of data, normally organized in some logical fashion so as to facilitate access (3.4). Such data may consist of factual information, statistics, numbers, textual, or composite records to be used as a basis for reasoning, discussion, or calculation. An entity within a data file is known as a *(data) record*. A collection of data files is sometimes known as a *databank*, particularly when the data files are electronically encoded (1.1.6).

Although data files may be encoded in any media (for example, a paper card index file is an example of a data file), the term has most often come to be used in connection with data files that are electronically encoded and stored in digital electronic form (3.3.1.6).

13 In keeping with the spirit noted in the Foreword that this Glossary is intended to be comprehensive but not exhaustive.

1.2.10.1 Table

A data file arranged into two-dimensional form, normally consisting of rows and columns together with headings or labels to depict the contents of the rows and columns. Tables may themselves contain other tables as elements resulting in a "latticed" arrangement of data. A *spreadsheet* is a special form of table originally used for accounting purposes and containing financial data, but which now includes a wide variety of complex reports arranged in tabular form, often with the aid of computer workstations (3.6.2.6).



1.3. Document Periodicity

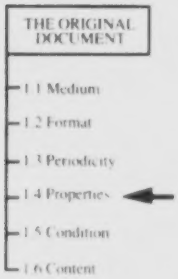
***Periodicity* refers to the number of parts into which the document is divided and the manner or sequence in which those parts are or have been published.**

1.3.1. Monograph

A *Monograph* is a published work, collection, or other document that is not a serial (1.3.2).

1.3.2. Serial

A *Serial* is a publication issued in successive parts, bearing numerical or chronological designations, at regular or irregular intervals and intended to continue indefinitely.



1.4. Document Properties

***Document Properties* refers to a classification of various components of documents as to their different tonal or color content and as to the types of objects[14] they contain. Emphasis is placed on those properties most closely associated with documents produced on paper.**

1.4.1. Tone

Tone refers to the color quality or color content of the document or parts of the document regardless of form or material content.

1.4.1.1 Monotone

Monotone documents (or parts of documents) are printed or otherwise produced using one color hue[15] only, most often black or near-black.

1.4.1.1.1 Two-Tone

Those parts of a monotone document that are represented in only two contrasting tones (regardless of the hue of the color, although the term is most often associated with black hues), with no intermediate shades. Thus, for purposes of this Glossary, a book printed with red ink on yellow paper would be considered two-tone. When one of the shades is black or near-black, and the other white or near-white, the document is described as being produced in *black-and-white*.

1.4.1.1.2 Greyscale

Those parts of a monotone document that are presented using a range of tones (regardless of the hue of the underlying color). The range of tones may either be *continuous* (such as in a

14 The Term "object" is used here in a sense that is more familiar to computer professionals than to librarians.

15 Strictly speaking, monotone documents should be termed "monohue".

photograph), where all possible values may essentially be taken on, or *discrete*, where only a finite set of values may be taken on.

1.4.1.2 Highlight Color

A two-tone (1.4.1.1.1) document, parts of which additionally contain areas highlighted with a second single color of uniform shade.

1.4.1.3 Two Color

A document containing two colors, intermixed to create intervening hues, and two extreme tones (normally black and white) used to create a continuous or discrete (see 1.4.1.1.2) range of shades.

1.4.1.4 Full Color

A document containing or attempting to contain a full range of colors, normally of all hues, tones, and shades.

1.4.2. Object Type

Object Type (see also Footnote 13) is a descriptor that conveys information about a given sub-area (*object*) of the document with regard to the manner in which it conveys data or information.

1.4.2.1 Text Objects

Text Objects are document objects consisting of written or printed (or otherwise displayed) stored words or ideograms.

1.4.2.2 Data Objects

Data Objects are document objects consisting of factual information normally arranged into datafiles (1.2.10) or tables (1.2.10.1) which are used as a basis for reasoning, discussion, or calculation.

1.4.2.2.3 Table

See 1.2.10.1.

1.4.2.3 Graphic Objects

Graphic Objects are document objects containing image information consisting of artwork, photographs, technical drawings etc, perhaps containing limited amounts of text usually as captions or for labelling purposes.

1.4.2.3.1 Line Art

Graphic objects created entirely from the use of text, dots, and straight or curved lines.

1.4.2.3.1.1 Graphs

Line art objects consisting of representations of the interrelationships of data in pictorial form.

1.4.2.3.2 Halftone

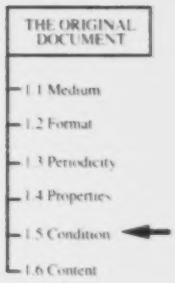
A representation of a greyscale (1.4.1.1.2) or color graphic object as a series of dots obtained, for example, by photographing or scanning an image through a mesh screen. By limiting the dots to, say, black and white (for example, by using high-contrast film), the illusion of greyscale may be created in a two-tone or black-and-white document (1.4.1.1.1).

1.4.2.3.3 Discrete Tone

A greyscale or color (1.4.1.4) graphic object where the tones take on discrete (normally equispaced) values within a range.

1.4.2.3.4 Continuous Tone

A greyscale (1.4.1.1.2) or color (1.4.1.4) graphic object where the tones fall continuously across an entire range of values, such as in a photograph (1.1.4, 1.2.9.3).



1.5. Document Condition

***Condition* refers to the physical state of the document compared with its state when originally published. The following presents only those characteristics of the physical state of a document that**

are pertinent to the main thrust of this Glossary, that is, to the paper milieu.

1.5.1. Archival

A document that can be expected to be kept permanently as closely as possible to its original form. An *archival document medium* is one that can be "expected" to retain permanently its original characteristics (such expectations may or may not prove to be realized in actual practice). A document published in such a medium is of *archival quality* and can be expected to resist deterioration.

Permanent paper is manufactured to resist chemical action so as to retard the effects of aging as determined by precise technical specifications. *Durability* refers to certain lasting qualities with respect to folding and tear resistance.

See also 3.3.5.

1.5.2. Non-Archival

A document that is not intended or cannot be expected to be kept permanently, and that may therefore be created or published on a medium (1.1) that cannot be expected to retain its original characteristics and resist deterioration.

1.5.3. Acidic

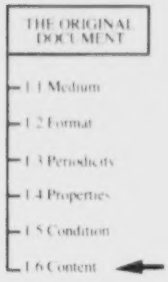
A condition in which the concentration of hydrogen ions in an aqueous solution exceeds that of the hydroxyl ions. In paper, the strength of the acid denotes the state of deterioration that, if not chemically reversed (3.1.2), will result in embrittlement (1.5.4). Discoloration of the paper (for example, *yellowing*) may be an early sign of deterioration in paper.

1.5.4. Brittle

That property of a material that causes it to break or crack when depressed by bending. In paper, evidence of deterioration usually is exhibited by the paper's inability to withstand one or two (different standards are used) double corner folds. A *corner fold* is characterized by bending the corner of a page completely over on itself, and a *double corner fold* consists of repeating the action twice.

1.5.5. Other

There are many other conditions that characterize the condition of a document. Bindings of books, for example, may have deteriorated for a variety of conditions. Non-paper documents may exhibit a variety of conditions (see, for example, 3.3.5 for a discussion of the concept of "Useful Life"). However, with the focus on paper original documents and on media conversion technologies for preservation, a full analysis of document condition would be beyond the scope of this Glossary.



1.6. Document Content

***Document Content* refers to the substance of the *material* or *information* within the document that is intended to be communicated.**

1.6.1. Intellectual Content

Intellectual Content refers to the ideas, thought processes, artistic expressions, etc., contained within the document.

1.6.2. Copyright[16]

Copyright refers to a means of legal protection provided to the author(s) of original published and unpublished works that have been "fixed in a tangible form of expression," in order to afford such authors the exclusive right of *exploitation*, in particular the right to control the reproduction, distribution, performance, or display of the work, or to control the

16 Copyright law as it applies to the subject of preservation will be the subject of a forthcoming paper by the Commission on Preservation and Access.

preparation of derivative works.[17] Often, exploitation of the work by others requires the consent of the author(s) and the payment of a *royalty* to the author(s), usually in the form of a fixed sum of money for each copy made, shown, or distributed.

For works copyrighted in the United States after January 1, 1978, protection afforded to the author(s) or the author(s)' estate is usually for the author(s)' lifetime plus 50 years. For works created prior to that date, the copyright period was 28 years from the date of publication (or the date of registration of copyright for unpublished works), plus an additional period of 47 years for works whose copyright was renewed during the last year of the first term.

Works published in the United States may be afforded protection in countries that were members of the Universal Copyright Convention or of the Berne Convention for the Protection of Literary and Artistic Works. Conversely, works published in such member countries are protected within the United States.

Most works that are the subject of preservation interest were published before 1978. The copyrights on the majority of those works were not renewed for the optional second term. Thus, the copyrights have expired on most of the works of current preservation interest that were subject to United States copyright protection. However, since this is not true of all such works, the normal practice is to check copyright ownership to verify clearance.

1.6.3. Structure

Structure refers to the divisions within a document provided for ease of access, reference, and other purposes. The broad structure of a given document is likely to vary according to its format (1.2), and there is also not necessarily any standard structure for a given format. With its long history, the structure of the printed book (1.2.2) has evolved towards a somewhat standard structure. Because of the focus of this Glossary on the preservation of the printed book, a typical book structure is presented here and structures for other formats are omitted.

17 For a fuller explanation of copyright laws, see "Copyright Basics", Circular No. 1, published by the Copyright Office of the U.S. Library of Congress, Washington, DC 20559.

1.6.3.1 Abstract (see 3.4.1.2)

1.6.3.2 Title Page

The *Title Page* of a work normally contains the title of the work, its author(s), and the name of the publisher.

1.6.3.3 Table of Contents (see 3.4.1.3)

1.6.3.4 List of Figures, Tables, Maps or Other Illustrations (see 3.4.1.4)

1.6.3.5 Preface (see 3.4.1.5)

1.6.3.6 Introduction (see 3.4.1.6)

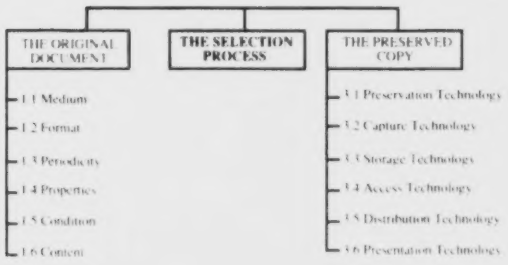
1.6.3.7 Body

The *Body* of a document refers to the main corpus of the work. It may be divided into *chapters, papers, articles,* or other segments.

1.6.3.8 Index (see 3.4.1.7)

1.6.3.9 Other

This category includes publisher's notes, credits, frontispieces, and other minutiae of publication.



2. THE SELECTION PROCESS[18]

The *Selection Process* refers to the means whereby original documents are selected for preservation purposes. The choice of selection strategy may be intrinsically affected by the choice of preservation or media conversion technology used (see 3.1), since the latter may well affect costs and other parameters associated with the former. Thus, the total costs of preservation will be a complex combination of the effects of selection strategy and choice of technology.

Thus, for example, with the use of microform (3.1.4), it is highly desirable (if not imperative) to obtain a complete copy of the document to be preserved prior to recording. This may require replacing missing or damaged pages from the prime copy being microfilmed, and the expense of obtaining these pages from copies held in other libraries. Microfilming also places a premium on recording only once. With the use of digital technologies (3.1.5), on the other hand, such replacement pages could be scanned at a later date and electronically "edited" into the main electronic document: with digital technologies, it may in fact be cheaper to scan more than one copy to facilitate such "editing" rather than to expend excessive manual labor on assembling the most perfect paper copy possible prior to microfilming.

The following is a brief — and very over-simplified — classification of selection methodologies. It is only intended to sketch the range of possibilities and not to do full justice to the complexity of this subject. It merely indicates some of the main lines of strategy or process used in selecting documents for preservation. Furthermore, often a combination of approaches is used rather than any single approach, with the actual condition of the document being the dominant factor in the choice.

18 See also "Selection for Preservation of Research Library Materials," a Report of the Commission on Preservation and Access, August 1989.

In all cases, the "universe" of documents to which the selection strategies outlined in this Section are applied is those documents that are deteriorating or are likely to deteriorate, such as brittle books or, more generally, books printed on acidic paper. "Preservation", however, may also be applied to the conversion onto other media of materials that, while in quite good condition, are scarce or unique, thus allowing patrons to handle facsimiles instead of the precious originals.

The term "essentially all documents" is used below to define documents from within the former universe that fit within the indicated selection strategy, while allowing that a number of these selected documents may yet be rejected following review for various reasons (such as having deteriorated to the point that preservation is not possible, or because it has been determined that the document has already been preserved elsewhere).

2.1. By Title

Selection is made from among individual works, perhaps by professional bibliographers who, possibly working in consultation with others, make a determination of the value of the selected work to a given collection, discipline, or field of study.

2.2. By Category

Selection is made by choosing essentially all documents from a within a given category, such as within a given time period, or of a given format (for example, all newspapers), subject classification, special collection, or , say, American imprint. The essence of this approach is that all documents within the category be readily and conveniently definable and accessible, without having to resort to time-consuming selection processes.

Colloquially, this approach is sometimes erroneously termed the *"vacuum cleaner approach"*, an appellation that is overly pejorative insofar as some prior review is almost always made to reject materials within a category that for various reasons are not suitable or desirable for preservation. In particular, a check is made to ensure that the material has not already been preserved.

Selection, for example, by time period permits the focus of effort on those periods of highest risk of deterioration with respect to paper-manufacturing processes.

2.3 **By Bibliography**

Selection is made by choosing essentially all documents specified in a published bibliography.

2.4. **By Use**

Selection is made by choosing essentially all documents in poor condition that are actually used by patrons as judged by some criterion such as, for example, frequency of circulation.

2.5. **By Condition**

Selection is made by preserving the documents in the worst physical condition.

The foregoing are examples of selection according to certain established *criteria*. Selection may also be made according to established *procedures*:

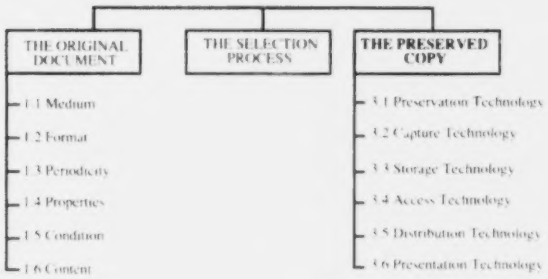
2.6. **By Scholarly Advisory Committee**

Selection is made with the assistance of a committee of scholars knowledgeable in a particular field who choose the material they consider to be of most importance to that field.

2.7. **By Conspectus**

Selection is made from institutional collections determined in a program initiated by the Research Libraries Group (RLG)[19] and described in the RLG Conspectus. The Conspectus describes collections on various levels from Level 0 (Out-of-Scope, a level which is in fact non-existent), through Level 4 (Research), to Level 5 (Comprehensive). Collection development officers (selectors) in about 50 major research libraries in the U.S. have evaluated their own collections to provide such brief descriptions. The Conspectus can be used as one of several means to determine "Great Collections."

19 The Research Libraries Group, Inc., is a not-for-profit corporation owned and operated by its governing members: major universities and research institutions in the United States.



3. THE PRESERVED COPY

This section addresses technologies employed in the preservation process. The first section broadly classifies different kinds of preservation processes. The remaining sections focus on the different technological stages associated with preservation processes dependent upon media conversion technologies. These are: capture technologies, storage technologies, access technologies, distribution technologies, and presentation technologies.

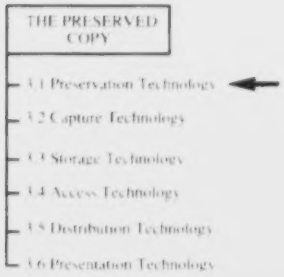
The divisions among these various stages of technology may, at first, seem artificial, particularly to those used to working with paper. For example, we distinguish between the storage medium (3.3.1), the distribution medium (3.5.1), and the presentation medium (3.6.1). In the world of paper, as stated in the Introduction, these are usually all one and the same, even though the same paper book, say, may play different roles at different times. When it is on the library bookshelf, it is a *storage* medium; when it is being messengered through inter-library loan, it is the *distribution* medium; and when it is being read by the patron, it is the *presentation* medium. In the world of convertible technologies, the separation becomes more than convenient sophistry — it becomes essential, since different media may well be used at any stage of the process. Consider, for example, a table from a scientific journal article (paper: the storage medium), which is FAXed across the nation using a data network (digital electronic: the distribution medium), and printed out directly onto photographic slides (film; the presentation medium) for projection in a lecture.

Indeed, in the preservation milieu, this conceptual separation also offers considerable flexibility. It offers the flexibility of separating the act of preservation itself from the ultimate means of storage and delivery. Thus, for example, microfilming may be used as a preservation process (3.1.4), but the microfilm contents may be printed later onto paper for user presentation purposes. Or the microfilm may be digitally scanned

and the contents stored on computer files for subsequent distribution across networks. As another example of this flexibility, images scanned and stored using digital preservation techniques (3.1.5) may later be interpreted using internal character recognition (3.2.5) or page recognition (3.2.6) technologies.

The point is that the ultimate use of the preserved document may not be well-articulated at the time of preservation. Thus, preservation technologies that offer the greatest flexibility are to be preferred to those (such as photocopying (3.1.3)) that offer less flexibility, although lack of funds and patron preference often dictates the use of the latter.

The distinction between the various technology stages is maintained throughout this Glossary.



3.1. Preservation and Media Conversion Technologies

Many different technologies have been proposed to address the problems of preservation. These can be divided into three broad categories: those directed at preserving both the content and physical embodiment of the original, those directed at preserving the content and copying the physical embodiment, and those directed at preserving the content only, without concern for the physical embodiment. Conservation and paper deacidification fall into the first category. The remaining technologies described below fall into the other categories.

In the second category every effort is made to copy the physical embodiment or format of the original as faithfully as possible, normally onto another medium. The term *media conversion* technologies is thus used for this class (*note*: this does not exclude copying a paper document onto another paper document: media conversion has still occurred). Media Conversion includes photocopying (3.1.3), microform recording (3.1.4), and the use of electronic digitization techniques (3.1.5).

The third category makes no attempt to preserve or copy the physical embodiment of the original. For example, merely rekeying the text (see 3.2.8) of a document composed entirely of text preserves only content and nothing else if no attempt is made to capture font and other formatting information.

Among librarians, the term "reformatting" has traditionally been used for "media conversion." The former term is not used in this Glossary because of possible confusion with the concept of Document Format (1.2). Furthermore, "reformatting" does not do justice to the concept of copying onto microform (3.1.4) or of digital scanning (3.1.5).[20]

This necessarily brief glossary of different preservation approaches also summarizes some of the key issues involved in comparing the various alternatives.

3.1.1. Conservation Treatment[21]

The treatment of a document to preserve it in its original form, in recognition that the original medium, format, and content are all important for research and other purposes. Pure conservation approaches are normally hand-tailored to the individual document and, as such, may be relatively expensive. Use is normally, therefore, limited to those situations where such expensive treatment is justified by the research requirements.

3.1.2. Paper Deacidification and Strengthening[22]

The treatment by chemicals to stabilize a document (in paper, by alkalization to neutralize the acid content) and/or to strengthen it (in paper by the use of a support coating or by impregnation). The alkalization treatment also usually entails depositing an alkaline reserve to buffer against further acidification.

Deacidification or strengthening can be applied to individual documents or, with some treatment processes, to a large number

20 It is tempting to use the term "remediate" for "media conversion," a temptation that has been resisted in the formulation of this Glossary.

21 For a discussion of the importance of conservation see "On the Preservation of Books and Documents in Original Form," by Barclay Ogden, Report of the Commission on Preservation and Access, October, 1989.

22 For more information see "Technical Considerations in Choosing Mass Deacidification Processes," by Peter G. Sparks published by the Commission on Preservation and Access, May 1990.

of documents at once (*mass* or *bulk deacidification*). The latter is a relatively cheap approach, and pilot plants have been or are being established in a number of countries to support different processes. There is, however, no standard approach at this time even though there appear to be a number of promising alternatives. There are also a number of unanswered questions at this time regarding the longevity of chemical stabilization processes, toxicity, the feasibility of scaling processes to full production requirements, the potential continuing "offgassing" implications to patrons resulting from the storage of thousands of treated volumes in confined library spaces, and other issues. Recent research appears to be addressing many of these concerns.

Deacidification is essentially a stabilization process that arrests deterioration. It does not turn brittle books back to their original state, although coating or impregnation can strengthen the paper to extend its useful life. Its greatest utility may lie in arresting embrittlement in books that are not too far gone, or for prophylactic protection of new or old books that have not yet started to turn brittle. Deacidification may also "buy time" in anticipation of later preservation by other processes.

3.1.3. Photocopying

Photocopying refers to the process of preserving the document by making a full-size (usually bound similarly to the original) facsimile copy on archival (1.5.1) paper by creating a photographic copy of the images of the pages contained in the document, possibly using a *photocopier* (3.2.1). As used here, photocopying refers to an in-line process where the original is scanned and one or more photocopies made all in one pass, with no form of retained intermediate storage being automatically generated (as contrasted with *microform recording* (3.1.4)) so that more copies can be made in the future. In actual practice, however, when photocopying is used for preservation it is customary to make a second photocopy that is retained in unbound form, so that further copies can readily be made in the future from this master copy.

A distinction is made between straight photocopying, which does not necessarily involve the use of archival paper (1.5.1), and *preservation photocopying*, which does require the use of archival paper.

The advantages of making such a facsimile are that normally a single paper facsimile is produced that is quite faithful to the

original, there is no machine interface required other than the photocopier itself, the medium (1.1) and format(1.2) of the original are retained, and the cost is usually less than other processes, particularly if the original is a monochrome document. Furthermore, library patrons prefer paper facsimiles to the use of, say, microforms (3.1.4), except where bulky documents, such as newspapers, are involved. The disadvantages, as compared with microform recording (3.1.4) and electronic digital preservation (3.1.5), is that normally second copies made from the master copy are of poorer quality than, say, prints of microforms made from master microforms. Furthermore, the costs of making subsequent copies is higher than the cost of printing microforms. Another disadvantage, shared to a greater or lesser extent with microforms, is that photocopying does not precisely reproduce all the information in the original, and there is some loss of information, especially for graphic objects (1.4.2.3) involving other than line art (1.4.2.3.1).

3.1.4. Microform Recording

Microform Recording refers to the process of preserving the document by filming the original document onto a microform film negative (1.1.2), that is, storing microimages of the pages or segments of the document on film. Positive film copies, which can be produced inexpensively, are made from this original film negative or *master*. Such a positive copy is both a storage (3.3) and distribution (3.5) technology, and is normally viewed using a *microform reader* (3.6.2.2), or paper positive prints may be made from the positive microform using printing devices designed for the purpose. Access to microfilm (1.1.2) using such a reader is serial (cf 3.3.1.6), whereas access to microfiche (1.1.2) is random (cf 3.3.1.6) like a book.

The advantages of microform are that the process is economically competitive with other processes; that film has a long useful life (3.3.5); and that microform copies — made from a second negative[23] (known as the *printing master*) copied from the original negative — may be made cheaply and distributed among other institutions, so that access is not limited to a single facsimile. Microform preservation is a well-tried, tested, and accepted method of preservation.

23 The original, or preservation, negative should not be viewed with a microform reader (3.6.2.2) because of potential damage to the negative.

The disadvantages are that there is usually a loss of information in the recording process, particularly in recording continuous tone imagery (1.4.2.3.4), since the film used is usually of high contrast;[24] and that readers dislike using microform readers compared with, say, reading books.

Mircoform-preserved documents can subsequently be converted to other media besides paper. They can be scanned (3.2.3) and converted to digitally-encoded documents (3.1.5) to take advantage of the benefits of digital encoding for storage, distribution, and access. However, any loss of information in the original recording process will be perpetuated in the subsequent digital recording.

3.1.5. Electronic Digitization

Electronic Digitization refers to the capture of the document in electronic form through a process of scanning (see 3.2.3) and digitization. The scanned image is stored electronically, usually on magnetic (see 3.3.1.6.1 and 3.3.1.6.2) or optical (see 3.3.1.6.3 and 3.3.1.6.4) storage media. The electronically stored image may be further *transformed* for reasons such as compression (see 3.3.2) or information interpretation (see 3.3.3); and subsequently *selected* through the use of access technologies (see 3.4), *distributed* through the use of distribution technologies (see 3.5), or *viewed* through the use of presentation technologies (see 3.6).

When originally scanned, or as a result of subsequent transformations, the document may in whole or in part be stored in *image (3.1.5.1), unformatted text (3.1.5.2.1), formatted text (3.1.5.2.2), or compound (3.1.5.3)* form. The distinction is important insofar as it affects *inter alia* the extent to which information such as text in the scanned document may be interpreted (3.2.5, 3.2.6, 3.2.7) and used for purposes of information access (3.4, in particular 3.4.2, but see also 3.1.5.1, 3.1.5.2, 3.2.4). An *image* representation is an electronic pictorial representation composed of dots (black and white, greyscale, or color) much like a halftone (1.4.2.3.2) printed photograph, and no distinction is made between text and other information (such as graphs, pictures, and so forth) contained in the document — in other words, the letter "b" is not stored as a character *per se*, but as a *"digital picture"* of the letter "b", and the series of numbers stored to represent the picture would be quite distinct

24 Newer processes becoming available appear to remove the obstacle of high-contrast recording.

among different typestyles used. *Text* representations, on the other hand, represent text as text, with a specific code used to denote the letter "b" independent of what typestyle is used.

Image representations cannot be searched for words or phrases: text representations can. Image representations of text may be converted into formatted or unformatted text representations using OCR (3.2.4) or ICR (3.2.5) techniques, but with loss of accuracy. In the context of preservation, image representations are likely to dominate, since the cost of transforming image into text representations with sufficient accuracy may be prohibitively high, at least in the immediate future. Thus full-text searching, for example, is not likely to be a feature of digitally-preserved documents. This is unlike the situation that exists with documents where the text already exists in digital electronic form, such as if the publisher had preserved the original tapes used in typesetting.

If and when OCR techniques are able to convert image format to text format with sufficient accuracy and performance, then the archives of digitally-preserved material in image format can be converted to text format using ICR (3.2.5) techniques, provided the original material was scanned with sufficiently high resolution (3.2.3). Furthermore, promising research has been done recently on the searching of documents for retrieval purposes using the "corrupted" (erroneous) text derived from the OCR or ICR scanning of image documents at existing levels of OCR/ICR accuracy and performance.

The advantage of electronic digitization is that it potentially combines the advantages of photocopying and microform recording while eliminating some of the disadvantages. Paper facsimiles can be produced at will by *printing-on-demand* (3.5.4) on paper (or writing the appropriate signals on whatever might be the appropriate output medium, in the case of video, film, or sound), thus eliminating the need for awkward microform readers. Alternatively, the stored images can be reconstructed and viewed at computer workstations (3.6.2.6). Furthermore, the stored digital images can be distributed essentially at will across data networks (3.5.5) for sharing among institutions. The content of the stored images can also be interpreted at any time (3.2.5, 3.2.6, 3.2.7) after recording (whenever it might become economically desirable to do so) for purposes of, say, creating indices for access purposes (3.4.1).

Another key advantage is the robustness of digital encoding. Further copies, including copies made in new formats (3.3.3) on other digital electronic storage media (3.3.1.6) for purposes of extending the useful life of the digital copy (see Introduction and 3.3.5), can be made without loss of information, as contrasted with photocopying (3.1.3) or microform recording (3.1.4). Furthermore, scanned images can be digitally enhanced (3.2.9) to improve the image quality.

The disadvantages are that this is a new and relatively untried technology, and the cost and other trade-offs are uncertain at this time. There are also concerns about the useful life (3.3.5) of present storage media, both in terms of the physical properties of the media and in terms of the robustness of the recording format (3.3.3) and of the means of access. Some, however, take the view that it will be both functionally and economically imperative in any event to recopy the data from storage medium to storage medium every few years to take advantage of the rapidly declining storage costs and increasing storage capacities of the technology, and that the useful life of a given medium is not the relevant issue (see Introduction and 3.3.5).

3.1.5.1 Image Document

A representation of the document *image* is electronically captured (usually with the aid of a digital image scanner — see 3.2.3) or created without interpretation of its actual *content*. This is stored as a sequence of 1's or 0's (known as *bits*), a "digital photograph" as it were. In certain image representations, a "1" indicates "black" and a "0" indicates "white" (*Binary Encoding*), but usually the representation is encoded in more complex representations (see 3.3.4 Encoding Method). In some representations, for example, the average grey level of a small area of the page, termed a "pixel", is encoded (*Greyscale Encoding*. See also 1.4.1.1.2). Such a pixel is a grey dot. The number of dots per inch is termed the *pixel resolution*. This pixel resolution may range from 100 per inch to several thousand per inch.

It is not unusual, for reasons of storage economy, to convert a greyscale-encoded image document into a binary-encoded image document of higher resolution at the time an image document is stored. Compression techniques (3.3.2) are used to achieve this. The resultant stored image represents a compromise between scanning resolution, image fidelity, and storage space.

The electronically-encoded sequence of 1's and 0's that represent an Image Document is also known as a *Bitmap*.

Image Documents are generally accessed by associating an index entry, such as a page number, with a segment of the Image Document. See

discussion following under 3.1.5.2 regarding other issues associated with searching and retrieving Image Documents.

3.1.5.2 Text Document

The text of the document only is captured as *character* representations, that is, each alphabetic character has a unique representation (see discussion above) following a standard means of encoding, such as the *ASCII* standard. With electronic digital storage, the amount of space taken to store a *representation* of a character generally takes far less than the amount of space taken to represent a character in *image* form. Usually, each character representation of a letter of, say, the Roman alphabet takes 8 bits (1 byte) of storage space. When stored in image form, the representation may take several orders of magnitude more storage space, depending upon the size of the character, the scanning resolution, and the degree of compression (see 3.3.2) used. See also 3.3.4.2.

Storing a document as a text document facilitates full-text or partial-text retrieval (see 3.4.2), where documents or parts of documents can be selected and retrieved by searching for the occurrence of keywords or strings of text. This is not possible with Image Documents (3.1.5.1), unless they have been wholly or partially converted to Text Documents using Optical Character Recognition (OCR) techniques (3.2.4, 3.2.5), a process that is not sufficiently accurate for most preservation purposes (see, however, 3.2.4 for a discussion of the use of such techniques for the construction of indices).

3.1.5.2.1 Unformatted Text

The character representation of the text contains no information to indicate font style, font size, or page layout. In this sense, unformatted character text representations are an example of irreversible compression (see 3.3.2.3).

3.1.5.2.2 Formatted Text

The character representation of the text also contains sufficient information to describe one or more of font type, font size, or page layout. In this sense, formatted text may, if the document segment contains only textual material, represent a form of reversible compression (see 3.3.2.2).

3.1.5.3 Compound Document

The document is captured as a combination of image and formatted or unformatted text.

3.1.6. Rekeying of Text

Rekeying of Text refers to a preservation technology where the text in a document is literally reentered by hand into a

composition or other device for republication or reproduction purposes, often with the use of a digital computer. See also 3.2.8.

3.1.6.1 Unformatted Text

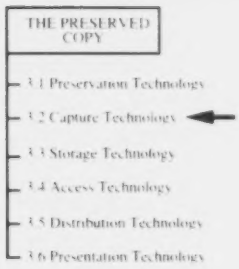
In the rekeying of the text, no attempt is made to key sufficient information to indicate font style, font size, or page layout.

3.1.6.2 Formatted Text

In the rekeying of text, information is captured to indicate one or more of font style, font size, or page layout.

3.1.7. Reprinting or Republication

The document is preserved by producing a new edition or reprint, possibly by reprinting from retained intermediate forms of the document, such as reprinting a book from photocomposition tapes. Alternatively, the document may be recreated from scratch.



3.2. Capture Technology

Capture Technology **refers to the technology used to transform the images or information contained in the original document into some other form, the form dependent upon the overall *media conversion* technology being used. This term is not relevant to Conservation (3.1.1) or Deacidification (3.1.2), which are *conservation* technologies, and do not employ media conversion techniques. Printing (see 1.1.1) on paper, is of course also a capture technology.**

3.2.1. Photocopier

A *Photocopier* is a device for making photographic copies of graphic images. A common form of the photocopier involves the use of the *xerographic* process, where light reflected from the original document is focused onto an electrically charged insulated photoconductor, and the latent image is developed using a resinous powder. For the purposes of this Glossary, the term *photocopier* is restricted to devices that use *analog* technologies, such as the use of light lens technology. *Digital* technologies are incorporated separately (see 3.2.3). With photocopiers so defined, the image is normally scanned and printed essentially in a single operation, and an intermediate scanned latent image is not normally stored for re-use at a later stage — although the two stage processes of photography, which indeed may be used for photocopying, do permit the use of the photographic negative as an intermediate storage device (a particular case of which is the use of microform recording technology — see 3.2.2).

3.2.2. Microform Recorder

A *Microform Recorder* is a camera or other photographic device for photographing the original document and printing it onto one of several forms of microform (1.1.2). The microform film in essence becomes both a storage medium (see 3.3.1.2) and a presentation medium (see 3.6.1.2 and 3.6.2.1). Other film copies and paper copies may also be made from the microform negatives for presentation (see 3.6.1.2).

3.2.3. Digital Image Scanner

A *Digital Image Scanner* is a device for scanning the images contained on pages of a document and transforming the scanned image into digital electronic signals corresponding to the physical state at each part of the search area, that is, into image documents (3.1.5.1). These signals are most often stored (see 3.3) for subsequent interpretation (see 3.2.5, 3.2.6, 3.2.7, and 3.3.2, 3.3.4), access (3.4), distribution (3.5), or presentation (3.6). A single small element of the document (known as a "pixel") is thus encoded quantitatively by a digital number, where the number contains sufficient information to represent the *image* content of the pixel (see 3.1.5.1). A digital image scanner on its own does not interpret the image information. The number of pixels per square inch is considered to be the *resolution* of the scanner. Typical resolutions with current technology range from 100

pixels per linear inch to over 1,000 pixels per linear inch, but there are trade-offs between resolution, speed, cost, and quality.

Digital Image Scanners may scan in one or more different modes, depending upon their capability and depending upon whether they are scanning monotone or color (1.4.1), or whether they are scanning line art, greyscale, halftone, or continuous tone objects (1.4.2.3, 3.1.5.1). Performance, in terms of speed, accuracy, and resolution depend upon the degree to which these attributes can be accommodated. The speed of digital image scanners range from one or two pages per minute to around fifty per minute.

A FAX machine (3.5.3) is a special form of digital image scanner. Other special forms of digital image scanners exist for scanning from media other than paper, such as digital image scanners that scan directly from microfilm (1.1.2). Such images scanned from microfilm, however, can be no better than the original microfilm image itself (see 3.1.4).

Digital image scanners may come equipped with different physical devices for accommodating the original documents. These may include flatbed platens equipped with manual feeds, semi-automatic feeds (one page at a time is fed into an automatic hopper), or fully-automatic feeds. Manual feeds offer the greatest safety from potential jamming, a point of importance in the scanning of unique documents. Flatbed scanners generally require either books to be disbound and one page at a time placed on the platen, or require books to be laid open face-down on the platen, which may cause some distortion. They may also come equipped with edge-scanners, which scan right up to the binding of the book, avoiding this distortion; or with cradle scanners, where the book is opened in a cradle (such devices are also used in some microform recording devices) and two angled scanning heads are lowered into the open, cradled book. In all cases, quality control of scanning is an issue with respect to fidelity of the scanned image and registration of the scanned image with respect to a defined standard.

3.2.4. Optical Character Recognition Scanner

An *Optical Character Recognition (OCR) Scanner* is a digital image scanner that in addition interprets the textual portion of the images and converts it to digital codes representing formatted or unformatted text (3.1.5.2). The less sophisticated such devices can only "recognize" one or a few fonts of a fixed

size, and can only interpret such information as unformatted text. The more sophisticated devices can represent multiple fonts of different sizes, and can interpret limited information as formatted text. At either extreme, no device achieves 100% recognition accuracy: accuracy of the better devices typically ranges between 95% and 98%, depending upon manufacturer imposed trade-offs between the sophistication of the device, its speed, and its intended range of applicability.

OCR devices are most often used where scanning errors and unformatted text are acceptable limitations, such as, for example, where the input material can be subsequently proofread and corrected, or where redundant information is scanned and the redundant information used to correct any inconsistencies arising from scanning errors (typically in certain commercial applications). In the context of document preservation, most uses of OCR devices are limited to where text information only suffices, and the form of the original document is not an important aspect of preservation. An important application is for use in the construction of indices for access and distribution (see 3.4 and 3.5), or for full contextual searching of information (3.4.2). Promising research has been done, for example, on the searching and retrieval of documents for retrieval purposes using the "corrupted" (erroneous) text derived from the OCR scanning of documents. The techniques utilized in this approach exploit the redundant information contained in the corrupted text.

Handwriting recognition devices, an extreme form of OCR devices, are not included in this Glossary. At this time, such devices are limited in capability.

3.2.5. Internal Character Recognition

Internal Character Recognition is the term sometimes used when the same interpretation technology that is used in OCR devices (3.2.4) is applied to an already stored digital image at a later date. This separates the functions of *scanning* the images (3.2.3) digitally, and of *interpreting* the images. Interpreting the scanned and stored images at a later date also allows for using different recognition technologies in the tradeoffs between accuracy, speed, and function. In the context of preservation and media conversion, it also allows for the immediate focus to be placed on scanning and storage (and possibly media conversion), deferring the option of character recognition and its applications (see 3.2.4) to a later date — at such time, massive-volume

character recognition and information interpretation is likely to be more economically feasible at higher levels of accuracy than with present technology.

3.2.6. Intelligent Character Recognition

Intelligent Character Recognition is the term sometimes given to Optical or Internal Character Recognition where the scanned and recognized information is further interpreted to take advantage of contextual information, that is, words, phrases, and so forth, rather than simply treating the text as a string of independent characters. Intelligent Character Recognition, for example, may be used by sophisticated computer programs to construct concordances automatically, or to create highly-sophisticated indexes. At this stage, intelligent character recognition is a field of research, rather than production, interest.

3.2.7. Page Recognition

Page Recognition is the term given to the automatic interpretation of features contained within the printed page such as titles, subheads, columns, paragraphs, figures, figure captions, footnotes, and so forth. Additional capabilities of sophisticated page recognition algorithms include the ability to determine fonts and font sizes. In essence, Page Recognition "reverse engineers" the image into marked-up copy.

3.2.8. Rekeying of Text

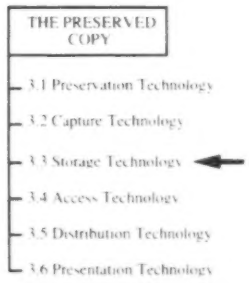
As an alternative or complement to OCR (3.2.4), textual information can be encoded by directly keying alpha-numeric text into computer files manually. This has some advantage in accuracy over OCR, but is slower. It may also be used in situations where the brittleness of acidic documents makes them so fragile that scanning technologies cannot safely be used. See also 3.1.6.

3.2.9. Enhancement

Enhancement refers to the use of mathematical algorithms to improve the quality of digitally scanned images (3.2.3), such as by computationally adjusting the contrast or brightness of the scanned image. The term also includes techniques that may be used to modify the scanned image for structural reasons, such as *bordering* to remove any unwanted scanned areas surrounding

the actual document pages, *de-skewing* to rectify the scanned image to correct for any skew in the placement of the document on the scanner, or *margin adjustment* to ensure that pages are properly aligned with each other.

A full glossary of terms associated with enhancement is beyond the scope of this document.



3.3. Storage Technology

***Storage Technology* refers to the technology used to store the images or information obtained through the use of some form of Capture Technology (3.2). This includes the *medium* used for storage (3.3.1), the *compression* methodology used to minimize the amount of storage medium employed (3.3.2), the *format* used to program the image or information onto the medium (3.3.3), the *encoding methods* used to represent any interpretation of the stored information (3.3.4), and the *useful life* of the storage medium (3.3.5).**

3.3.1. Storage Medium

3.3.1.1 Paper (see 1.1.1)
3.3.1.2 Microform (see 1.1.2)
3.3.1.3 Video (see 1.1.3)
3.3.1.4 Film (see 1.1.4)
3.3.1.5 Audio (see 1.1.5)
3.3.1.6 Digital Electronic

A family of storage devices where information or data are represented by a series of quantized changes to the surface of the storage medium, where such quanta are recorded or modified using electronic means. There are two main classes in this category: *magnetic* devices where, in recording, the magnetic state of a coated surface is altered by the electronic digital signal, and, in reading, the surface is sensed using reading heads conceptually similar to those used in common tape recorders; and *optical* devices where the optical properties of a coated

surface are altered (in one such technology, submicrometer-sized holes are recorded and read by laser beams focused by electronic means onto the area of the spot). The recorded quanta normally corresponds to a recorded "1" or a recorded "0", that is, of *bits (derived from "binary digits")*, all data and information being constructed from these basic building blocks.

Such devices are further classified according to whether they are *read/write* devices (that is, information may be written onto the device and read from the device, and the information can be modified as many times as desired), *read only memory (ROM)* devices (that is, prerecorded information can be read from the device, but the information cannot be modified), or *write-once-read-many (WORM)* devices (that is, information may be written once by the consumer onto the device, but thereafter it can only be read). Most optical devices are either read only or WORM devices, but a class of devices that combine both magnetic and optical technologies (*magneto-optical* devices) are indeed read/write devices.

Typically, magnetic devices are of higher performance in terms of *access time* to a given segment of recorded information and *transfer time* of such accessed information to the host device. Optical devices, however, are generally more economic in terms of storage capacity. Magnetic technologies have a longer history than optical technologies, and more is known about their useful life, for example (see 3.3.5). Both technologies seem to be following similar cost/performance curves with performance parameters doubling in capability approximately every two to three years (except for access times which are improving much more slowly), and cost per bit halving about every two to three years.

Both devices are further classified as to whether they are *random access* devices (such as *disk storage devices*) or *serial access* devices (such as *tape storage* devices). With random access devices, information stored at any point can be directly accessed (much as is accomplished by placing the playing-arm of a phonograph at any point on the phonograph record); with serial access devices, information can only be accessed by passing through information that may be recorded ahead of it on the medium (as in winding through a tape on a tape recorder to arrive at a particular passage).

3.3.1.6.1 Magnetic Disk

A rotating circular plate having a magnetized surface on which information may be stored as a pattern of polarized spots on concentric or spiral recording tracks. These plates or platters are usually stacked in *disk drives*, several to a drive. These platters may either be *removable* or not, although in high performance disk drives, the platters are usually not removable. They are, however, read/write devices (3.3.1.6). Some removable magnetic disks of lower capacity are known as *floppy disks*, since originally the recording medium was made of a flexible plastic.

3.3.1.6.2 Magnetic Tape

A plastic, paper, or metal tape that is coated or impregnated with magnetizable iron oxide particles on which information is stored as a pattern of polarized spots. These are read using magnetic tape drives. Access times with magnetic tapes are slower than those associated with correspondingly priced disks, since they are serial access devices, but the tapes are almost always removable so that the information can be stored *off-line*, thus making tapes[25] useful for archival storage (but see 3.3.5).

3.3.1.6.3 Optical Disk

A rotating circular plate on which information is stored as submicrometer-sized holes and is recorded and read by laser beams focused on the disk. This includes the class of *CD-ROM* devices, which embodies the same 5 1/4" diameter format used for CD recordings. CD-ROM's are usually read by inserting the CD-ROM disk into a *CD-ROM player*. Other typical formats involve 12" or 14" diameter formats, but there is a dearth of standards. The latter are usually read by inserting them into *optical jukebox* devices, which perform the role suggested by their name. Even when mounted, access times for optical disks are typically relatively slow, because of the lag time needed to "spin up" the disk. However, the cost per stored bit is extremely low. Error rates may also be higher than for magnetic technologies. As such, optical disks are most useful where there is an abundance of redundant information contained in the stored data, such as would be the case with the storage of scanned document pages. On viewing the data, the eye would not likely be troubled by a tiny dot among an ocean of dots being the wrong shade of grey. See also the discussion of magneto-optical devices (3.3.1.6.5). Conversely, magnetic devices excel in the recording of encoded text (see 3.3.4.2), but may be expensive to use for the storage of images even when compressed (3.3.2).

3.3.1.6.4 Optical Tape

An emerging class of technology that combines the advantages and disadvantages of tape (3.3.1.6.2) with those of optical recording technology (3.3.1.6.3). Their chief advantage may lie in very cheap cost per bit storage, but at this time they suffer from relatively high error rates.

25 Removable disks, such as floppy disks, are also used for archival storage. However, magnetic tapes are usually cheaper when large volumes of data are to be archived.

3.3.1.6.5 Magneto-Optical Disk

Disks that combine the use of magnetic and optical technologies. To record data, elements of the crystal structure of the substrate are aligned by using a laser to heat the element in the presence of an applied magnetic field. When the magnetic field is aligned one way, a "1" is recorded; when the magnetic field is reversed, a "0" is recorded. The data are read by reflecting a lower-intensity laser beam off the surface; the polarization of the reflected light varies according to the crystal alignment of the element of the substrate. Unlike regular optical disks, magneto-optical disks are read/write, and have performance characteristics somewhere between those of magnetic disks and optical disks in terms of access times, transfer rates, and storage capacity.

3.3.2. Compression

Compression refers to the extent to which the encoded form of the preserved or reformatted document has been modified to reduce the amount of storage space required by the storage medium. The technique takes advantage of the great redundancy that is present in much recorded data, particularly in image documents (3.1.5.1). Savings of storage of factors of ten or more may readily be achieved depending upon the scanning resolution and methodology employed (3.2.3), the type of material being scanned, and the particular compression method used. Although without compression the storage requirements grow rapidly as the square of the scanning resolution (3.2.3), with effective compression methods the storage requirements can be constrained to grow almost linearly with the scanning resolution. This is because advantage is taken of the greater data redundancy accruing from the increase of scanning resolution — compression effectively eliminates or reduces this data redundancy. Thus, the greater the redundancy of information contained in the scanned material, the more compression is possible — continuous tone photographs, for example, often contain large amounts of redundant information. Compression is an important factor in the economics and efficacy of digital preservation.

3.3.2.1 Uncompressed

No compression has occurred.

3.3.2.2 Reversibly Compressed

Compression has occurred so that the process can, if required, be reversed so that the original can be recovered without loss of information. Also known as "lossless".

3.3.2.2.1 CCITT Group Compression

Compression standards defined by the International Consultative Committee for Telephony and Telegraphy (Comité Consultative Internationale pour la Téléphonie et la Télégraphie).

3.3.2.2.2 Reversible Textual Compression

If sufficiently complete, the representation in whole or in part of documents as formatted text (3.1.5.2.2) may represent a form of reversible compression. The use of a markup language (3.3.4.3) is also a form of reversible textual compression. See also 3.3.4.

3.3.2.2.3 Page Description Language Compression (PDL)

See 3.3.4.4

3.3.2.2.4 Other Compression Standards or Algorithms

Refers to other compression standards, de facto standards, or algorithms.

3.3.2.3 Irreversibly Compressed

Compression has occurred so that the process cannot be precisely reversed. The original cannot be recovered without loss of information.

3.3.2.3.1 Irreversible Textual Compression

The representation in whole or in part of a document as unformatted or partially formatted text (3.1.5.2) may represent a form of irreversible compression. The content of the text may be obtained but not one or more of its font style, font size, or positioning on the page.

3.3.3. Storage Format

As used in information storage and retrieval, *Format* or *Storage Format* refers to the actual representation of the stored data on the storage medium, that is, the specific way in which it is encoded or programmed onto the medium. Classifying such methodologies is beyond the scope of this document. Indeed, for the most part — and particularly as applied to digital electronic

storage technologies — there are few general standards that are accepted by all or most manufacturers. The implication is that access to the information stored on the medium depends upon specific software or computer programs supplied by the manufacturer, software that may become obsolete with the passage of time. One result may be that stored information may need to be reformatted or transferred to newer storage media periodically in order for the information to remain accessible with current software and technology.

3.3.4. Encoding Method

Encoding Method refers to the *extent* to which the information *content* of the document has been interpreted and encoded, rather than merely recorded. Such interpretation may be beneficial for a number of reasons including as a means of achieving reversible compression (3.3.2.2); for the construction of document indices to facilitate searching and access (3.4.1); or for efficient distribution of the information across data networks (3.5.5). For example, a document that has been merely scanned as a bit-mapped image (3.1.5.1) has not been encoded (3.3.4.1), even though faithful "digital pictures" of the pages of the document have been obtained. If the images of the document text are later interpreted through internal character recognition (3.2.5), then the digital representation has been *textually encoded* (3.3.4.2).

3.3.4.1 No Encoding

No interpretation of the information contained in the original document has occurred. If the document were originally scanned using a digital image scanner (3.2.3), then the document in this instance is generally stored in some image format (3.1.5.1), compressed or not (3.3.2). If portions of the document were originally scanned using optical character recognition (3.2.4), then those portions will be stored as either formatted or unformatted text (3.1.5.2).

3.3.4.2 Textual Encoding

The text contained in the original document has been interpreted so that each character has a separate representation (see 3.1.5.2). Such interpretation may have occurred at the time of scanning if an optical character recognition device is used (3.2.4), or later using internal character recognition (3.2.5) programs applied to documents in image format (3.1.5.1). Such textual interpretation may result in either unformatted or formatted text, depending upon the degree of sophistication of the device or program. Recognition accuracy may also be limited.

3.3.4.3 Markup Language Encoding

A computer markup language is a means for describing, for an electronically stored document, the complete positioning, format, and style of text and image segment representations (3.1.5) within the document. When combined with textual representation, it is a means for achieving fully formatted text (3.1.5.2.1). When combined with relevant image information about document graphics material (if any), it may be a means of archiving fully reversible compression (3.3.2.2) of the document. An example of a markup language is *SGML (Standard Generalized Markup Language)* that has been adopted by the United States Government and by many publishers as a pseudo-standard.

3.3.4.4 Page Description Language Encoding

A computer language in which segments of text and images are economically described with respect to form, orientation, size, density, and other characteristics for purposes of economic transmission across networks and between host devices and output devices such as printers. Page Description Languages are another form of compression (3.3.2), as well as a form of encoding.

3.3.5. Useful Life

Useful Life refers to the archival quality of the storage medium. It usually refers to the period of time during which there is no unacceptable loss of information stored on the medium; and during which the storage medium remains usable for its intended purpose.

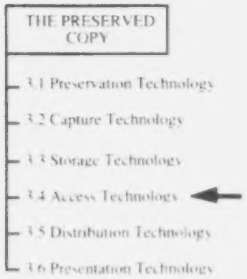
The longevity of paper varies considerably depending upon its method of manufacture and conditions of storage (see 1.5). Unless the paper is produced to meet permanent standards (1.5.1), paper may last from a few years or so to hundreds of years. Most paper produced since the middle of the nineteenth century has a useful life of less than 100 years. Paper produced to meet archival standards should last several hundred years. Film, provided it is manufactured, processed, and stored according to archival standards, appears to have a useful life well in excess of 500 years. Videotape appears to be extremely vulnerable and to have a relatively short life of a few decades.

Digital electronic storage media have a varying useful life projected to range from a few years to over 100 years. The latter has not been formally tested by experience, but is projected based on laboratory stress tests. Such media, however, become obsolete for other reasons long before their physical properties render

them useless (see, for example, 3.3.3). It becomes economically and functionally infeasible to maintain the information stored on the original medium of capture, since it becomes far cheaper to transfer the information periodically to higher density and cheaper newer technologies. Concerns also exist regarding the possibility of modifying digitally-encoded documents, particularly when "read/write" (3.3.1.6) devices are used (this is essentially not possible with "read only" or "write once, read many" technologies (3.3.1.6)); and regarding other issues of security.

The implications of periodic recopying for libraries are quite far-reaching. Libraries are not used to having to maintain their inventory by periodic recopying, even though such practices are quite common in data centers. Indeed, the recent impetus of preservation may have caused some librarians to rethink their position in this regard, although librarians still tend to think in terms of periods of centuries rather than having (or wanting) to recopy every few years. Such considerations may either hinder the adoption of digital technologies or eventually cause some rethinking of the underlying economics of librarianship.

Further implications are discussed in the Introduction.



3.4. Access Methodology or Technology

Access Methodology or Technology **refers to the means of selecting information from among all the information that is stored.**

3.4.1. Indexed Access

A *Document Index* is a systematically ordered file of objects[26] that refer to a collection of documents or to specific parts of those documents, organized in such a way as to facilitate searching the document collection for purposes of selection of single documents or groups of documents contained in the collection. Such document indices may be stored on different media depending upon how they are to be used.

3.4.1.1 Via Catalog

Access via a file of bibliographic records, created according to specific and uniform principles of construction and under the control of an *authority file*, which describes the documents contained in a collection. The file is usually organized in a systematic manner to facilitate access and document selection. Catalogs historically have been implemented in card files, but increasingly such card files are retroactively and prospectively giving way to computerized data files (1.2.10) which may be accessed and searched by patrons with the use of computer workstations (3.6.2.6) and data networks (3.5.5). Such computer-based catalogs are increasing in sophistication to support complex queries, including *Boolean* queries, which support logical searching (e.g., all the works of fiction written in Albania published between 1890 and 1919 by authors whose last name begins with the letter "L").

3.4.1.2 Via Abstract

Access via a summary of the document. Most often, the summary is of a contribution to a journal (1.2.6) or other periodical (1.3.2). Such a summary is usually without interpretation or criticism, and may contain a bibliographic reference (or *pointer*) to the original document. A collection of document abstracts may be used for purposes of search and selection (e.g., *Chemical Abstracts*, published by the American Chemical Society and also available in digital electronic form).

3.4.1.3 Via Table of Contents

Access via a list of parts contained in a document, such as chapter titles or articles in a periodical, with references by page number or other locator to the starting point of the particular part, usually ordered by sequenced groupings or the order of appearance. Collections of tables of contents may also be used for search and selection purposes.

26 See Footnote 13.

Other parts of documents that may be used for search and selection purposes include:

3.4.1.4 Via List of Figures, Tables, Maps or Other Illustrations

Access via a list of those parts of a document that are either figures, tables, maps or other illustrations, respectively, with location reference by page number or other locator, usually ordered by location of appearance within the document. Figures, tables, maps, etc. may be listed separately. Usually, in a document, these lists follow the Table of Contents in some order.

3.4.1.5 Via Preface

Access via a note preceding the body of a document that usually states the origin, purposes, and scope of the work(s) contained in the document and may include acknowledgements of assistance. When written by someone other than the author(s) of the document, the preface is more properly termed a *foreword*.

3.4.1.6 Via Introduction

Access via the material that heads the body of a document and that provides an overview of the work that follows, or other introductory material to the text.

3.4.1.7 Via Index

Access via a systematically ordered collection of words or other terms or objects[27] contained within a document, with references by page number or other locator to the placement of the object within the document for purposes of accessing the object. The index is usually placed last in a document.

3.4.1.8 Via Citation

Access via *reference* to a document or to a part of a document, such as an *article* in a journal (1.2.6). A *bibliography* is a collection of citations directed to a specific purpose, such as a *subject bibliography* or a bibliography of citations appended to a journal article.

3.4.2. Full (or Partial) Document Access

Full Document or *full text* searching is where the full text of a collection of documents is stored, and the entire text of all or portions of the documents is searched for specific character strings, usually combined with some Boolean logical searching

27 See Footnote 13.

capabilities. This requires that the document be textually encoded (3.3.4.2) either because it was initially created that way or perhaps more likely in the context of preservation because such textual encoding was obtained from scanned document images (3.1.5.1) with internal character recognition (3.2.5). Thus, for example, a search may consist of searching for all documents in the collection published by a given author or set of authors between certain dates containing the text "all that glitters." Full text searching is normally implemented on computers. For other than small collections of documents, a given search may be very costly in terms of computer processing time.

3.4.2.1 Via Inverted Text File Index

The use of *Inverted Text Files* (or other similar techniques) is often used as a compromise between indexed and full text searching. A file of words (*Keyword*), phrases (*Key Phrase*), or other text objects contained in a given collection of stored documents is created from an initial analysis of the full text together with locators as to where all instances of the word, phrase, or other object can be found within the file. In use, instead of the full text being searched for all occurrences of the object,[28] the inverted file itself efficiently gives pointers to the locations. The construction of such an inverted file, however, may be expensive for large collections of documents, as would adding new words or other objects[29] to the file at a later date. Furthermore, the use of the file is only as good as the care that has been given to the choice of objects to be contained within the file.

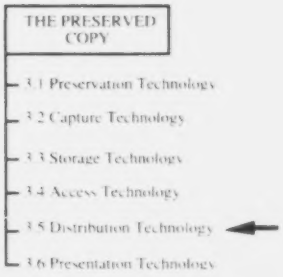
3.4.3. Compound Document Access

Compound documents are documents that contain both textually and other forms of encoded information, including image (see 3.3.4). Techniques are being developed for expanding the concept of text searching to searching of full compound documents, including those containing image objects[30]. A full glossary of such techniques, however, is premature and beyond the scope of this document.

28 See Footnote 13.

29 See Footnote 13.

30 See Footnote 13.



3.5. Distribution Technology

Distribution Technology **refers to the technology used to distribute or deliver the stored encoded document from one point to another. Some form of *delivery service* may be used (3.5.2), or, if the medium is paper, it may be distributed using point-to-point or distributed *FAX* (3.5.3). On the other hand, if the medium is digital electronic, then either the document may be converted to paper, by "*printing-on-demand*" (3.5.4) and subsequently distributed using delivery services or FAX, or *data networks* (3.5.5) may be used for distribution to a *computer workstation* (3.6.2) , possibly to be converted to another medium, such as paper, at the point of delivery (see 3.6.1).**

3.5.1. Distribution Medium

The *Distribution Medium* is the medium used to transport the stored encoded document to the presentation or viewing device (3.6.2). The same media that can be used for original documents (1.1) can also be used as distribution media.

3.5.1.1 Paper (see 1.1.1)
3.5.1.2 Microform (see 1.1.2)
3.5.1.3 Video (see 1.1.3)
3.5.1.4 Film (see 1.1.4)
3.5.1.5 Audio (see 1.1.5)
3.5.1.6 Digital Electronic (see 1.1.6)

Whichever technology is used for storage (3.3.1), digital technologies may usually be used as the medium of distribution, as contrasted with using delivery services (3.5.2) to deliver the document. Paper, for example, can be scanned and transmitted by FAX (3.5.3) or across data networks (3.5.5). The only exception to this at this time is video, which

is normally distributed by *analog* electronic distribution networks (as opposed to digital — see 1.1.6), because of the high information capacity (*bandwidth*) required. As the bandwidth of data networks grows, however, it is anticipated by many technologists that analog transmission will yield to digital transmission even for video recordings. Films, too, are often transmitted by converting them to video recordings (with some loss of quality at this time), and transmitting them across analog video networks.

3.5.2. Messenger Services

Messenger Services refers to the use of local, regional, or national messengering or mail services to hand-deliver documents from the point of inventory or storage to the patron or consumer. One special case of this includes the patrons performing the messengering services for themselves by viewing the document, or by directly acquiring it (purchasing or borrowing), at or from the location of the document's storage.

3.5.3. FAX

FAX or *Facsimile Transmission* is a system of communication or delivery for paper documents or other graphics material in which a special digital image scanner (3.2.3) scans the pages of the document, compresses the scanned image using CCITT Group Compression (3.3.2.2.1), and transmits the digital signals by wire or radio to a FAX receiver at a remote point. The FAX receiver decompresses the signals received and prints the digital image on paper. FAX transmission is a point-to-point protocol that is normally conducted over voice (3.5.6) or data (3.5.5) networks. Usually, scanning and printing devices are relatively slow (about 5 pages per minute), and the quality is limited. The popularity of FAX rests on its simplicity of use and the relatively low cost of the equipment. With the rapid growth of installed FAX equipment, FAX has recently been extensively used for inter-library loan purposes, and is also becoming used for intra-campus delivery purposes.

3.5.4. Print-on-Demand

Print-on-Demand refers to the capability to print documents right at the time they are required by patrons and consumers, rather than following traditional norms of printing documents in advance of need and coping with the need to distribute and inventory printed documents in anticipation of demand. This approach to distribution mirrors the "just-in-time" approach to inventory control. *Print-on-Demand* techniques are normally

used in conjunction with digitally stored documents (3.3.1.6) and data networks (3.5.5). The approach offers the promise of closing the gap between the world of digital technologies and those who maintain the superiority or simply prefer the characteristics of paper documents. Documents may be printed right in the patron's office or at a shared local facility from where it is delivered to or picked up by the patron.

3.5.5. Data Networks[31]

A *Data Network* is a communications network that transports data between and among computers and computer workstations (*network nodes*). Such networks may depend upon different physical media to transport the encoded digital signals (twisted pair copper wire, coaxial cable, fiber optic cable, satellite, and so forth); different protocols to encode the signals; and different ways in which the encoded signals are interpreted for use in applications. They also include bridges, routers, and gateways for connecting different media and for translating one protocol into another. Data networks vary considerably in speed and capacity, depending upon the physical media, the protocols used, and the particular architecture of the network. Network speeds and other performance characteristics appear to be more than doubling every two to three years.

3.5.5.1 Local Area Network

A *Local Area Network (LAN)* is a data network used to connect nodes that are geographically close, usually within the same building. In a wider view of a local area network, multiple local area networks are interconnected in a geographically compact area (such as a university campus), usually by attaching the LANs to a higher-speed local *backbone*.

3.5.5.2 Wide Area Network

A *Wide Area Network (WAN)* is a data network connecting large numbers of nodes and LANs that are geographically remote, such as within a broad metropolitan area, or between widely-separated metropolitan areas. This would also include *regional networks*, such as NYSERNet, which interconnects research and educational institutions in New York State.

31 The Technical Assessment Advisory Committee of the Commission for Preservation and Access is preparing a report on the implications of data networks.

3.5.5.3 National Network

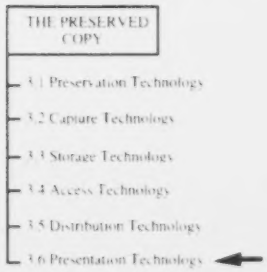
A WAN, or a federation of interconnected WANs, that span the nation, such as the NSFNet, BITNet, CSNet, CREN, and, more generally, the Internet and the anticipated NREN (National Research and Educational Network). These national networks often use a high-speed spanning national backbone to interconnect regional WANs. Protocols are established to facilitate routing of information across the national networks to users at connected nodes. The national networks often have international connections and outreach.

3.5.6. Voice Networks

Voice Networks are local, national, or international networks used to carry voice or telephone traffic. They may be either analog or digital (see 1.1.6). Because of different technical requirements, the transmission of data and voice usually is conducted using different transmission protocols, although it is increasingly common to share the same wiring plant. In general, there is increasing integration between the voice and data milieus.

3.5.7. Cable Networks

Cable Networks are local, regional, or national networks normally used for the transmission of analog (see 1.1.6) signals such as video (see 1.1.3) television signals.



3.6. Presentation Technology

***Presentation Technology* is the term given to technologies that present the encoded document to the end user or patron, possibly following some conversion of one medium to another. If the storage medium is paper, for example, no conversion would be necessary, and the storage medium and the presentation medium are one and the same (unless the**

distribution technology used were, say, FAX, in which case there are intervening conversion processes). If the storage medium, on the other hand, were digital electronic (3.3.1.6), for example, and data networks (3.5.5) were used as the means of distribution, then the presentation technology might be a computer workstation (3.6.2.4) or the distributed encoded document could be converted to some other form such as paper.

3.6.1. Presentation Medium

The *presentation medium* is the medium into which the stored document (3.3), which has been distributed over the distribution medium (3.5.1), is converted to facilitate viewing or reading by the end user.

3.6.1.1 Paper (see 1.1.1)
3.6.1.2 Microform (see 1.1.2)
3.6.1.3 Video (see 1.1.3)
3.6.1.4 Film (see 1.1.4)
3.6.1.5 Audio (see 1.1.5)
3.6.1.6 Digital Electronic (see 1.1.6)

3.6.2. Presentation or Viewing Device

A *Presentation or Viewing Device* converts the distribution medium (3.5.1) into the presentation medium (3.6.1). This includes the class of *computer workstations* (3.6.2.6).

3.6.2.1 Paper Document

A paper document, such as a book, must itself be considered a viewing device in this context when the presentation medium is paper (3.6.1.1). See 1.2 for a classification of different formats for paper documents.

3.6.2.2 Microform Reader

A display device with a built-in screen and magnification so that a microform (1.1.2) can be read comfortably at normal reading distances. Such devices may be accompanied by *microform printers* that can produce full-size (generally low-quality) paper copies of the microforms.

3.6.2.3 Video Projector (Television Set)

A device used to project or play back videotapes (1.1.3 and 3.6.1.3) onto a television screen. Normally this is accomplished through the use of a videorecorder (see below) and *television set* or *television projection system*. However, it is becoming increasingly common to play the video back through a computer workstation (3.6.2.6), possibly converting the analog signal to digital form (1.1.6).

The term *videorecorder* is often used to denote a device capable of both recording live television signals onto videotape and for reading recorded videotapes and transmitting the signal to a video projector or television set.

3.6.2.4 Film, Slide, or Other Projectors

A device to project motion picture films (1.1.4), still photographic slides (1.2.9.3), or other graphic materials (1.2.9) onto a screen, and, with some device, to reproduce sound from the film soundtrack. *Slide viewers* enable the user to view the slides through background projection on a small screen. Other classes of projectors (such as *overhead projectors*) are designed to project images recorded on transparencies onto a screen.

3.6.2.5 Audio Devices

A device capable of playing back audio documents (1.1.5) such as phonograph record players, CD players, and tape cassette players.

3.6.2.6 Computer Workstation

A device capable of supporting the creation, storage, access, distribution, or presentation of digital electronic documents (1.1.6), ranging from special purpose devices such as electronic typewriters through microcomputers to high-performance engineering or desktop publishing workstations or even large mainframe computers. They may vary considerably in performance, as typically measured by the computer's internal processing speed, storage capacity, and ability to move data between its various devices. The traditional distinction between a *personal computer (PC)* and a *high-performance workstation* is blurring, and the term *workstation* is generically used to cover both.

3.6.2.6.1 Display Monitor

That portion of a computer workstation used to view digital electronic documents. This may consist of a display module built into the computer or it may be physically separated from the computer, but attached by cable. Display monitors may be black-and-white (1.4.1.1.1), greyscale (1.4.1.1.2), or color (1.4.1.4). They may also come in varying physical sizes typically ranging from about 8″ on the diagonal to 23″ or more. They may also display with varying resolution, with the higher (but not highest) performance monitors capable of displaying over 1,000 x 1,000 pixels (spots).

3.6.2.6.2 Local Printer

A device locally attached to a computer workstation capable of printing digital electronic documents stored in the computer (3.3.1.6) or distributed to the computer from across a data network (3.5.5). Such devices may utilize a range of technologies including *impact printing*, *ink-jet printing*, *thermal printing* and *laser printing*. They may print at varying speeds ranging from 10 characters per second to some tens of pages per minute. They may print with resolutions varying from several dots per linear inch to several hundred dots per linear inch. They may print in black-and-white, greyscale, or color.

3.6.2.6.3 Remote Printer

A printer (3.6.2.6.2) that is accessible to a computer workstation remotely across a data network (3.5.1.6). These may typically be higher performance devices than local printers, particularly regarding speed or resolution. Such devices are typically shared among many uses and users. They may have special capabilities for "finishing" documents.

3.6.2.6.4 Other Local Media Output Devices

Computers capable of supporting multi-media (3.6.2.7) may support other "presentation" devices, such as television monitors for video recordings (although the trend is to combine the television video monitor and the computer display monitor into a single "head"), and audio playback devices for sound signals, including connections to "hi-fi" stereo equipment.

3.6.2.7 Multi-Media Workstation

A computer workstation (3.6.2.6) capable of supporting and combining multiple media such as digital electronic, video, sound, and paper.

4. SOURCES OF INFORMATION

Works referenced in the compilation of the Glossary include:

- The A.L.A. Glossary of Library and Information Science. Chicago: American Library Association, 1983.
- John Carter. A.B.C. for Book Collectors. New York: Alfred A. Knopf, 1980.
- John Dean. A Glossary of Library Technical Terms. Private Communication: October, 1989.
- Geoffrey Ashall Glaister. Glaister's Glossary of the Book. Berkeley: University of California Press, 1979.
- Nancy E. Gwinn (editor). Preservation Microfilming: A guide for Librarians and Archivists. Chicago: American Library Association, 1987.
- Dennis Longley and Michael Shein. Dictionary of Information Technology, Second Edition. Oxford University Press, New York, 1986.
- Ray Prytherch. Harrod's Librarians' Glossary, Fifth Edition. Grower Publishing Company, Brockfield, Vermont, 1984.
- Matt J. Roberts and Don Etherington. Book Binding and the Conservation of Books: A Dictionary of Descriptive Terminology: Washington: Library of Congress, 1982.
- McGraw-Hill. Dictionary of Scientific and Technical Terms. Fourth Edition, 1989.
- Rosenberg, Jerry M. A Dictionary of Computers, Data Processing, and Telecommunications. John Wiley and Sons, 1983.
- Bohdon S. Wynar. Introduction to Cataloging and Classification. Littleton, Colorado: Libraries Unlimited, Inc., 1985.
- Webster's New Collegiate Dictionary. G. & C. Merriam Co., 1979.

INDEX

Abstract 24, 51
Access 29
Access Technologies
  Compound Document Access 53
  Full Document Access 52
  Indexed 51
  Partial Document Access 52
Access Time 44
Acid Content 31
Acidic 21
Active Documents 3, 4
Alkalization 31
Analog 39
Architectural Plan 16
Archival 21
Archival Quality 21
Arrangement 14
Art Original 16
Audio 12
  Compact Disk 12
  Digital Acoustic Tape (DAT) 12
  Distribution Medium 54
  Document Medium 12
  Phonograph Disk 12
  Presentation Medium 58
  Record 12
  Storage Medium 43
  Tape cassette 12
Audio Device 59
Authority File 51
Bandwidth 55
Batch Processing 3
Berne Convention 23
Bibliography 52
  Subject 52
Binary Encoding 36
Bit 12, 36, 44
Bitmap 36
BITNet 57
Body 24
Book 15
Boolean Search 51
Bordering 42
Brightness Adjustment 42
Brittle 21
Buffering 31
Bulk Deacidification 32
Cable Networks 57
Capture 29
Capture Technologies
  Digital Image Scanner 39
  Enhancement 42
  Intelligent Character Recognition 42
  Internal Character Recognition 41
  Microform Recorder 39
  Optical Character Recognition Scanner 40
  Page Recognition 42
  Photocopier 39
  Rekeying of Text 42
Cartographic Materials 15
Catalog 51
  Authority File 51
CCITT 55
CCITT Group Compression 47
CD-ROM 45
  Player 45
Chemical Treatment 31
Citation 52
Coaxial Cable 56
Color 19
  Full Color 19
  Highlight 19
  Two Color 19
COM 11
  Computer-Output-On-Microfiche 11
Compact Disk 12
Compound Document 37
Compound Document Access 53
Compression 46
  CCITT Group 47
  Page Description Language 47
  Reversible 47
  Reversible Textual 47
  Uncompressed 46
Computer Workstation 58, 59
  Display Monitor 59
  Local Media Output 60
  Local Printer 60
  Personal Computer 59
  Remote Printer 60
Computer-Output-On-Microfiche (COM) 11
Condition 20
Conservation 30, 31
Conservation Technology 30
Content 22
  Copyright 22
  Intellectual 22

Structure 23
Continuous Processing 3
Continuous Tone 20
Contrast Adjustment 42
Copyright 22
Berne Convention 23
Universal Copyright Convention 23
Corner Fold 21
Double 21
CREN 57
CSNet 57
Data File 16
Databank 16
Record 16
Table 17
Data Network 56
Local Area Network (LAN) 56
National Network 57
Wide Area Network (WAN) 56
Data Object 19
De-Skewing 43
Deacidification 31
Bulk 32
Mass 32
Depreciation 3
Digital 12, 39
Digital Acoustic Tape (DAT) 12
Digital Electronic 12
Distribution Medium 54
Document Medium 12
Presentation Medium 58
Storage Medium 43
Digital Electronic Storage 43
Access Time 44
CD-ROM 45
Floppy Disk 44
Magnetic 43
Magnetic Disk 44
Magnetic Tape 45
Magneto-Optical 44
Magneto-Optical Disk 46
Optical 43
Optical Disk 45
Optical Jukebox 45
Optical Tape 45
Random Access 44
Read Only Memory (ROM) 44
Read/Write 44
Serial Access 44
Write Once Read Many (WORM) 44
Digital Image Scanner 39
FAX Machine 40
Digitization 34
Discrete Tone 20
Disk 44, 45, 46
Magnetic 13
Magneto-Optical 13
Optical 13
Disk Drive 44, 45, 46
Display Monitor 59
Distribution 29
Distribution Medium 54
Audio 54
Digital Electronic 54
Film 54
Microform 54
Paper 54
Video 54
Distribution Technologies
Cable Networks 57
Data Network 56
Distribution Medium 54
FAX 55
Messenger Services 55
Print-On-Demand 55
Voice Networks 57
Document 9
Arrangement 14
Condition 20
Content 9, 22
Definition 9
Document Index 51
Format 2, 9, 14
Layout 14
Medium 2, 9, 10
Original 9
Properties 18
Style 14
Type 14
Document Access
Full 52
Partial 52
Document Condition 20
Acidic 21
Archival 21
Brittle 21
Non-Archival 21
Yellowing 21
Document Index 51
Document Medium 10
Audio 12
Digital Electronic 12
Film 12
Microform 11
Multi-Media 13
Paper 10
Video 11
Document Periodicity 17

Monograph 17
Document Properties 18
Object Type 19
Tone 18
Documents
Active 3, 4
Double Corner Fold 21
Drawing 10
Durability 21
Electronic Digitization 34
Bitmap 36
Compound Document 37
Image Document 36
Text Document 37
Encoding Method 48
Markup Language 49
None 48
Page Description Language (PDL) 49
Textual 48
Enhancement 42
Bordering 42
Brightness Adjustment 42
Contrast Adjustment 42
De-Skewing 43
Margin Adjustment 43
Facsimile 25
Facsimile Transmission 55
FAX 55
FAX Machine 40
Fiber Optic Cable 56
Film 12
Distribution Medium 54
Document Medium 12
Photograph 12
Presentation Medium 58
Storage Medium 43
Film Projector 59
Filmstrip 16
Floppy Disk 44
Foreword 52
Format 14
Book 15
Cartographic Materials 15
Data File 16
Graphic Materials 16
Manuscript 15
Music 16
Newspaper 15
Pamphlet 15
Periodical 15
Printed Sheet 15
Formatted text 37
Formatted Text Rekeying 38
Full Color 19
Full Document Access 52
Key Phrase 53
Keyword 53
Via Inverted Text File Index 53
Full Text Searching 35, 52
Glossary 5
Scope 5
Structure 7
Graph 20
Graphic Materials 16
Graphic Object 19
Continuous Tone 20
Discrete Tone 20
Halftone 20
Line Art 20
Graph 20
Great Collections 27
Greyscale 18
Greyscale Encoding 36
Halftone 20
Handwriting 10
Handwriting Recognition 41
Hard Copy 11
Highlight Color 19
ICR 35, 42
Image 39
Image Document 36
Binary Encoding 36
Bitmap 36
Greyscale Encoding 36
Impregnation 31
Index 24, 52
Via Abstract 51
Via Catalog 51
Via Foreword 52
Indexed Access 51
Via Citation 52
Via Index 52
Via Introduction 52
Via List of Figures 52
Via Preface 52
Via Table of Contents 51
Instance 3
Intellectual Content 22
Intelligent Character Recognition 42
Internal Character Recognition 41
Introduction 1, 24, 52
Inverted Text File Access 53
Irreversible Compression 47
Textual 47
Irreversible Textual Compression 47
Journal 15
Jukebox, Optical 45
Key Phrase 53

Keyword 53
LAN 56
Layout 14
Lifecycle Costing 3
Light Lens Technology 39
Line Art 20
List of Figures 52
List of Figures Etc. 24
Local Area Network (LAN) 56
Local Printer 60
Longevity 49
Magazine 15
Magnetic Disk 13, 44
Magnetic Storage Devices 43
Magnetic Tape 13, 45
Magneto-Optical Disk 13, 46
Magneto-Optical Storage 44
Manuscript 15
Margin Adjustment 43
Markup Language
  SGML 49
Markup Language Encoding 49
Mass Deacidification 32
Media conversion 25, 30
Media Conversion Technology 30
Medium 10
Messenger Services 55
Microfiche 11
Microfilm 11, 33
Microform 11
  Distribution Medium 54
  Document Medium 11
  Presentation Medium 58
  Storage Medium 43
Microform Reader 33, 58
Microform Recorder 39
Microform Recording 33
  Copies 33
  Master Negative 33
  Printing Master 33
Microimage 11
Monograph 17
Monotone 18
Multi-Media 13
Multi-Media Computer Workstation 60
Music 16
National Network 57
Newspaper 15
Non-Archival 21
NREN 57
NSFNet 57
Object 19
  Data Object 19
  Graphic Object 19
  Text Object 19
Object Type 19
OCR 35
OCR Scanner 40
Off-Line Storage 45
Optical Character Recognition Scanner 40
Optical Disk 13, 45
Optical Jukebox 45
Optical Storage Devices 43
Optical Tape 13, 45
Original Document 9
Page Description Language (PDL) 47, 49
Page Recognition 42
Pamphlet 15
Paper 10
  Distribution Medium 54
  Document Medium 10
  Drawing 10
  Handwritten 10
  Hard Copy 11
  Parchment 11
  Presentation Medium 58
  Printed 10
  Storage Medium 43
  Strengthening 31
Paper Document 58
Parchment 11
Partial Document Access 52
PC 59
PDL 49
Periodical 15
  Journal 15
  Magazine 15
Periodicity 17
Permanent 21
Personal Computer 59
Phonograph Disk 12
Photocopier 32, 39
Photocopying 32
  Preservation 32
Photograph 12, 16
Picture 16
Pixel 36, 39
  Resolution 36
Platen 40
Postcard 16
Poster 16
Preface 24, 52
Presentation 29
Presentation Device 58
Presentation Medium 58
  Audio 58
  Digital Electronic 58
  Film 58

Microform 58
Paper 58
Video 58
Presentation Technologies
Presentation Device 58
Presentation Medium 58
Viewing Device 58
Preservation 25
Preservation Technology 30
Chemical Treatment 31
Conservation 31
Deacidification 31
Electronic Digitization 34
Microform Recording 33
Paper Strengthening 31
Photocopying 32
Rekeying of Text 37
Reprinting 38
Republication 38
Preserved Copy **29**
Print-On-Demand 35, 55
Printed Sheet 15
Printer 60
Printing 10, 60
Chemical Transfer 10
Impact 60
Ink-Jet 60
Laser 60
Photographic 10
Thermal 60
Thermographic 10
Typographic 10
Xerographic 10
Printing Master 33
Prints 16
Processing
Batch 3
Continuous 3
Properties 18
Random Access 44
Read Only Memory (ROM) Storage 44
Read/Write Storage 44
Record 12
Reference 52
Reformatting 30, 31
Rekeying of Text 37, 42
Formatted 38
Unformatted 38
Remote Printer 60
Reprinting 38
Reproduction 16
Republication 38
Research Libraries Group 27
Resolution 39
Reversible Compression 47
Reversible Textual Compression 47
RLG 27
ROM 44
Satellite Transmission 56
Scanner 39
Platen 40
Scanner Resolution 39
Scanning 34
Scope of Glossary 5
Selection 25
By Bibliography 27
By Category 26
By Condition 27
By Conspectus 27
By Scholarly Advisory Committee 27
By Title 26
By Use 27
Great Collections 27
Vacuum Cleaner Approach 26
Selection Process **25**
Serial 17
Serial Access 44
SGML 49
Slide 16
Projector 59
Viewer 59
Spreadsheet 17
Stereograph 16
Storage 29
Storage Format 47
Encoding Method 48
Storage Medium 43
Audio 43
Digital Electronic 43
Film 43
Microform 43
Paper 43
Video 43
Storage Technologies 43
Compression 46
Format 47
Medium 43
Useful Life 49
Structure 23
Abstract 24
Body 24
Index 24
Introduction 24
List of Figures Etc 24
Other 24
Preface 24
Table of Contents 24
title Page 24

Structure of Glossary 7
Style 14
Table 17, 19
Table of Contents 24, 51
Talking Book 9, 14
Tape 45
  Magnetic 13
  Optical 13
Tape Cassette 12
Tape Drive 45
Technical Drawing 16
Technologies
  Access 29
  Capture 29
  Digital 12
  Distribution 29
  Presentation 29
  Storage 29, 43
Technology 29
  Media Conversion 30
  Preservation 30
Telephone Systems 57
Television Set 58
Text
  Formatted 37
  Unformatted 37
Text Document 37
Text Object 19
Text Searching 35
Textual Encoding 48
Title Page 24
Tone 18
  Black-and-White 18
  Greyscale 18
  Highlight Color 19
  Monotone 18
  Two-Tone 18
Transparency 16
Twisted Pair Copper 56
Two Color 19
Two-Tone 18
Type of Document 14
Uncompressed 46
Unformatted Text 37
Unformatted Text Rekeying 38
Universal Copyright Convention 23
Useful Life 49
Vacuum Cleaner Approach 26
Video 11
  Distribution Medium 54
  Document Medium 11
  Presentation Medium 58
  Storage Medium 43
  Television Playback 11
  Videodisk 11
  Videotape 11
Video Projector 58
Videorecorder 59
Viewing 29
Viewing Device 58
  Audio Device 59
  Computer Workstation 59
  Microform Reader 58
  Multi-Media Workstation 60
  Paper Document 58
  Projector
    Film 59
    Slide 59
    Video 58
  Television Set 58
  Videorecorder 59
Voice Networks 57
WAN 56
Wide Area Network (WAN) 56
Workstation 58, 59
WORM 44
Write Once Read Many (WORM) Storage 44
Xerography 39
Yellowing 21